

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2002, 02:13:10 ; Search time 2452.42 Seconds

(without alignments)
10947.858 Million cell updates/sec

Title: US-09-758-017A-1

Perfect score: 1283

Sequence: 1 gacatccgcgtgcaaatatg.....agaataataattatattta 1283

Scoring table: IDENTITY_NUC

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_jom.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hlg_hum.*
31: em_hlg_inv.*
32: em_hlg_other.*
33: em_hlg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	1278.2	99.6	1283	5	GM0275971	AJ275971	Gadus mor
2	1164.2	90.7	1355	5	GM0275972	AJ275972	Gadus mor
3	364.2	28.4	1857	10	MMU55041	U55041	Mus musculi
4	362.6	28.3	2002	10	BC011039	BC011039	Mus muscu
5	361	28.1	1875	10	MMU08975	X99018	M. musculus
6	361	28.1	1951	10	MMU08975	X98975	M. musculus
7	355.4	27.7	2045	9	BC015205	BC015205	Homo sapi
8	355.4	27.7	2053	6	A64377	A64377	Sequence 1
9	355.4	27.7	2057	6	HS090008	Y09008	H. sapiens m
10	355.4	27.7	2062	9	HSUNG	X15653	Human cDNA
11	355.4	27.7	2122	6	AX281824	AX281824	Sequence
12	319	24.9	1737	10	BC004037	BC004037	Mus muscu
13	262.8	20.5	189672	9	CNS01DXY	AL136039	Human chr
14	262.8	20.5	213353	2	AC010258	AC010258	Homo sapi
15	261.2	20.4	123159	2	AC091880	AC091880	Homo sapi
16	261.2	20.4	171125	9	AC025437	AC025437	Homo sapi
17	226.8	17.7	130932	9	AC092658	AC092658	Homo sapi
18	210.8	16.4	12337	1	AE005488	AE005488	Escherich
19	210	16.4	21722	1	AE008820	AE008820	Salmonell
20	210	16.4	274050	1	AL627276	AL627276	Salmonell
21	207.6	16.2	1532	1	ECOUNG	J03725	E. coli ung
22	207.6	16.2	1532	6	AX277028	AX277028	Sequence
23	207.6	16.2	10272	1	ECORBA	D13169	E. coli gene
24	207.6	16.2	12510	1	AE000344	AE000344	Escherich
25	207.6	16.2	16446	1	D90886	D90886	E. coli geno
26	207.6	16.2	29254	1	ECCK12RIII	D64044	Escherichia
27	206.8	15.7	270365	1	AJ414153	AJ414153	Yersinia
28	201.4	15.5	12918	1	AE001926	AE001926	Delonococ
29	199.2	15.4	78736	2	AC095886	AC095886	Rattus no
30	198.2	14.5	10937	1	AE004306	AE004306	Vibrio ch
31	186.2	14.0	657	6	AX083978	AX083978	Sequence
32	180	14.0	11847	1	U32687	U32687	Haemophilu
33	180	13.9	158325	2	AC087564	AC087564	Homo sapi
34	178.6	13.4	10144	1	AE002470	AE002470	Homo sapi
35	172.2	13.4	191549	9	AC010582	AC010582	Homo sapi
36	172.2	13.4	199420	9	CNS01DXY	AL136040	Human chr
37	172.2	13.4	349980	6	AX044031	AX044031	Sequence
38	172.2	13.4	349980	6	AX044032	AX044032	Sequence
39	170.8	13.3	42071	1	SC2661	AL359949	Streptomy
40	170.8	13.3	42071	1	SC2661	AE006042	Pasteurel
41	169	13.2	12736	1	AE006042	AL591792	Sinorhizo
42	168.8	13.2	333800	1	SME591792	AE009193	Agrobacte
43	165	12.9	12258	1	AE008159	AE008159	Agrobacte
44	165	12.9	13567	1	AE008159	AL162755	Neisseria
45	161	12.5	331801	1	NMA422491		

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
1	GM0275971	Gadus morhua mRNA for uracil-DNA glycosylase precursor (ung1 gene).	AJ275971.1	GI:7413911	precursor; ung1 gene; uracil-DNA glycosylase.	Atlantic cod.	Gadus morhua	1283 bp	Lanes O., and Willassen, N.P.	Identification, cloning and expression of nuclear and mitochondrial uracil-DNA glycosylase from Atlantic cod(Gadus morhua).	Submitted (02-MAR-2000) Lanes O., Department of Biotechnology, University of Tromsøe, Institute of Medical Biology, Medical
2	GM0275971	Gadus morhua mRNA for uracil-DNA glycosylase precursor (ung1 gene).	AJ275971.1	GI:7413911	precursor; ung1 gene; uracil-DNA glycosylase.	Atlantic cod.	Gadus morhua	1283 bp	Lanes O., and Willassen, N.P.	Identification, cloning and expression of nuclear and mitochondrial uracil-DNA glycosylase from Atlantic cod(Gadus morhua).	Submitted (02-MAR-2000) Lanes O., Department of Biotechnology, University of Tromsøe, Institute of Medical Biology, Medical

FEATURES Facility, N-9037, NORWAY
source Location/Qualifiers
1..1283
/organism="Gadus morhua"
/db_xref="taxon:8049"
/tissue_type="liver"
/dev_stage="adult"
/country="Norway"
18..923
/gene="ung1"
18..923
/gene="ung1"
/EC_number="3.2.2.3"
/product="uracil-DNA glycosylase precursor"
/protein_id="CAB85707.1"
/db_xref="GI:7413912"
translacion="MLFKILGICORCISNRVLPGLIPOTLCFSKLMITPKLRSSN
VEOKTSPOLSYEOLERAKNKKALDKIRAKATPAGGEMPRRLAEFEKPYKOL
MSFYADERSRHTVYPADOVYSWTMCIDIVKVIILGODPHGNQAHGICFSYQK
VPPPSLVNIXKEICTDIDGFKHPGHGDLGSAWAKGVLLNLNVLIVRAHQASHKDRG
WETFDVAIKMLSVNREGVLELLMGSYAHKKGATIDRRNHVLAQVHPSPLSAHNGFL
GCKHPSKANGLIKLSGTEPIWMRAL"
BASE COUNT 364 a 279 c 280 g 360 t
ORIGIN

Query Match 99.6%; Score 1278.2; DB 5; Length 1283;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1280; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gacatcgccttgcaaatatggttgtaaggttaaggttaagcaagaatgcaatacaca 60
DB 1 GACATCGGCTTCAAAATATGTTTCAAATAGGGTTATGCCAAAGATCATATCATCA 60
QY 61 atcgggtgttaccaaggtttactaatccccaacttatgttttctaataaagaaga 120
DB 61 ATCGGGTCTTACCAAGTTTACTAATTTCCCAACTTATGTTTCTTAATTAATCAAGA 120
QY 121 taacgcgaagaactgaggttccctcaaatgttgaaacaaagacgtccacagctt 180
DB 121 TAACGCCGAAGAACTGAGGTCCTCAATGTGGAACAAAAGCGTCATGCCACACTTT 180
QY 181 cagtgagcagctgtgaagaatgagccaaataagaagaagcagcgttgacaagaat 240
DB 181 CAGTGAGCAGCTGGAATAATGGCCAAAATAAGAAACAGCGCTTGACAAATTAAGAG 240
QY 241 caaaagcaagcgtgcaggtttccgagagacttgagagaagagagcgtgcagagtt 300
DB 241 CAAAAGCAACGCTGCAGGTTTGGAGAGACTTGGAGAGAGAGCTGGCTGCAGAGTTTG 300
QY 301 aaaagccatacttcaaaacatltgalttcccttgaactgtatgagagagccgtcacac 360
DB 301 AAAAGCATACTTCAAAACATGATGCTCTTGTACTGATGAGAGAGCGCGTCAACCG 360
QY 361 tctaccaccagcgtgatcaagtgtacagttcgacagagatgtgtgacattcaagaat 420
DB 361 TCTACCCACCGGCTGATCAAGTGTACAGTTGGACAGAGATGTGTGACATTCAGAGCTGA 420
QY 421 aagtgatgatttcaagcagagcccttaccagagttcccaagaagacatgagcttgt 480
DB 421 AAGTACTGATTTCTAGGCGGAGACCTTACACAGGTTCCCAACCAAGCACATGGACTGTGT 480
QY 481 tcaagtgtcaaaagcaggttcccccctcccccagttctgtgaacatatacaaaagatt 540
DB 481 TCAGTGTGCAAAAGCAGTTCGCCCTCCGCCAGCTCTGTAACAATCAAGAATTTGT 540
QY 541 gtaccgacaatgtatggtctcaagcatctctgacatgagatcttaagcggatgggcaaac 600
DB 541 GTACCGACATTTGATGGCTTCAAGCATCTTGACATGAGATCTAAGCGATGGCAAAAC 600
QY 601 aaggagtgctgtcttaacgcggtgtgacagttgaggccataagggccaactcccaaa 660
DB 601 AAGGAGTGTCTGTCTTAACGCGGTGTGACAGCTGGGGCCCATCAGGCCAACTCCCAAA 660

DB 601 AAGGGTGTCTGCTTAAACGGGCTGTGACCGTGGGGCCCATCAGGCCAACTCCCAAA 660
QY 661 aggacagagcctgagagacccctcaccagcgtgtgatacaagtgtgagcgtcaaacgg 720
DB 661 AGGACAGAGGCTGGAGACCTTCACCGACGCTGTGATCAAGTGTGGCTGACGTCACCGGG 720
QY 721 aaggatgcttttccctgt 780
DB 721 AAGGATGCTTTTCTGTTGTGGGCTCATACCCCATGAAGAGGAGGACCATCGACA 780
QY 781 ggaacgtcaccaatgtctgcaagcgtgtcatcctccttctgtctgtcatcgttgg 840
DB 781 GGAACGTCAACCATGTCTTGCAAGCTGTTCATCATCTCTTGTGTGTCATCGTGGGT 840
QY 841 tctgtgtgtgaagcattctccaaagcctaaagcgtgtgtgtgtgtgtgtgtgtgtgt 900
DB 841 TCTGTGTGTGAAGCATTCTCCAAAGCTAAACGCGCTGTGAACATATCTGGGACGAGC 900
QY 901 ctataactggagagacctaactcttattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 960
DB 901 CTATAACTGGAGAGACCTTAACCTTTATGCTGCTTATACTGTTAAGTGTTAAGA 960
QY 961 tgacatacaactataatttctcaagcgtttccaaagtccaacaaactataaagcttca 1020
DB 961 TGAACATCAACATATTTTCTACAGCTTTTCCAACTTCAACCAATCATATAAGCTTCA 1020
QY 1021 ttgtcttttggaatgagcgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
DB 1021 TTGTCTTTTGGAAATGATCTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
QY 1081 tctgcagctgtgactgt 1140
DB 1081 TCTGCCATGTGTGCTCATGTGTGCAATATTAACCTTTCACAACTTAACAAATATTTAT 1140
QY 1141 ttataatgtatataatctgtacataaagattgttttccagcagctgttctcaag 1200
DB 1141 TTATATATGATATATTTCTGTATTAAGATGTTTTTCCAGCGGTTCATTAAG 1200
QY 1201 tactagataactaaactgttataactatlttcaatgatgtaactgtcttaagtttca 1260
DB 1201 TACTAGATATTAACCTGTATTAACCTTATTTCCATGATGTCMACTGCTTAAGTTTAA 1260
QY 1261 tgcagaataaattatataatta 1283
DB 1261 TGCAGAAATTAATTAATTAATTA 1283

RESULT 2
GMO275972 1355 bp mRNA linear VRT 04-APR-2000
LOCUS Gadus morhua mRNA for uracil-DNA glycosylase precursor (ung2 gene).
DEFINITION AJ275972
VERSION AJ275972.1 GI:7413913
KEYWORDS precursor; ung2 gene; uracil-DNA glycosylase.
SOURCE Atlantic cod.
ORGANISM Gadus morhua
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
REFERENCE 1 (bases 1 to 1355)
AUTHORS Lanes, O. and Willassen, N.P.
TITLE Identification, cloning and expression of nuclear and mitochondrial uracil-DNA glycosylase from Atlantic cod(Gadus morhua).
JOURNAL Characterisation of the cold-active catalytic domain Unpublished
AUTHORS Lanes, O.
REFERENCE 2 (bases 1 to 1355)
AUTHORS Direct Submission
TITLE Submitted (02-MAR-2000) Lanes O., Department of Biotechnology,
JOURNAL University of Tromsøe, Institute of Medical Biology, Medical
FEATURES Facility, N-9037, NORWAY
source Location/Qualifiers
1..1355

```
/organism="Gadus morhua"
/db_xref="taxon:8049"
/tissue_type="liver"
/dev_stage="adult"
/country="Norway"
90..995
/gene="ung2"
90..995
/gene="ung2"
/EC_number="3.2.2.3"
/codon_start=1
/product="uracil-DNA glycosylase"
/protein_id="CAB85708.1"
/db_xref="GI:7413914"
/translation="MIGQOHINSEFSPVSKRKRVSKELEKTEKHAEEVOITPKLRSSN
VEOKTSSPOLSEVLEERMAKKKALDKIRKATPAGFETWRELAEPKPYKOL
MSFVADERSRHTVYPADQVYSPTEKCDIOQVYVILIDDPYHGNQAHGCFSTQK
VPPPSLVNITKELCTDIDGRKPHGSLSNAROGVLLNNAVILTRHONSHDRG
WETPDVNIKLSVNRSEVFLMGSYAHKKGATIDRRHHVLAQVHPSPLSAHNGFL
GCKHPSKANGLKLSGTEPINMRAL"
BASE COUNT      393 a      287 c      304 g      371 t
ORIGIN
Query Match      90.7%: Score 1164.2; DB 5; Length 1355;
Best Local Similarity 98.5%: Pred. No. 1.7e-286;
Matches 1175; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
OY 91 aaacttattgtttcttaattatgaataacgcggaagaactgagtcctcaatg 150
DB 163 AACCGAAGAGATGCGGAGAGAGTTCAAGATACCGCGAAGAACGAGTCTCTCAATG 222
OY 151 tggaaacaaagacgtcatcgccacagcttcagtggaagcagctggaagaatggccaaa 210
DB 223 TCGAACAAAGACGTCATGCGCACAGCTTTCAGTGAGCAGCTGGAAGAATGGCCAAA 282
OY 211 ataaagaagcagcgtcttgaagattagagcaaaagcaagcgtcaggttcogagaaga 270
DB 283 ATAAAGAAAGACCGGCTGACAAAGATTAGACAAAGCAAGCGCTGAGGTTTCGGAGAGA 342
OY 271 ctggaagaagaagcgtgctgagagttgaaagccacttcaacaacttgatgctc 330
DB 343 CTTGGAGAGAGAGTGGCTGCAGAGTTTGAAAGCGCATACTTCAACATTTGATGCTCT 402
OY 331 ttgtgctgtaagaagggcgctcacacccgtctaccacggcgtgataagtgtacgt 390
DB 403 TTGTAGCTGATGAGAGAGACCGCTCACACCGCTACCCAGCTGATCAAGTGTACAGTT 462
OY 391 cgacagaagtggtgacattcaagaatgtgaagaatgatttcttagcgccagacccttacc 450
DB 463 GGAACAGAGATGTGTACATTCAAGATGTAAAGTAGTATTCTTAGGCCAGACCCCTTACC 522
OY 451 acggtcccaacaaagacatggaactctgttccagtgcaaaagcaggtcccccctccc 510
DB 523 ACGGTGCCAAGCAACAGACATGACCTCTGTTTCAGTGTGAAAAGCCAGTTCCCCCTCCC 582
OY 511 ccagtcctgtaacatatcaagaatgtgtacacgaacttgatggcttcaagcatcctg 570
DB 583 CCAGTCTCTGTAACATATTCAAAGAAATGCTGTACCCACATTTGATGGCTTCAAGCATCTG 642
OY 571 gacatggagaatcgaatggatggcaaaacaaagggtgtgtctgtcttaacgggtgtccta 630
DB 643 GACATGGAGATCTAAAGCGGATGGCAAAACAAAGGGGTGCTGTGCTTAAACGGCGGTCTTA 702
OY 631 ccgtgcggggcccatcaggccaactcccaagaagacaggtctggagagacttcaaccgacg 690
DB 703 CCGTGGGGCCCATCAGCGCAACTCCACAAAGACAGAGAGTGGGAGACTTCAACGAGAG 762
OY 691 ctgtatcaatgtgctgagcgttcaacggggaagagctgttctcctgtgtgtgggctcat 750
DB 763 CTGTGATCAAGTGCTGACGCTCAACCGGGAAGGAGTGTGTTTCCGTTGTGGGCTCAT 822
OY 751 acgccataaagaaggagagaccatcgacgaaggaacgltacacatgtcttgcaagctgttc 810
```

```
|||||
DB 823 ACGCCATTAAGAGAGGAGGACCATCGACAGGAAAGTACCATGCTTGAAGCTGTTCC 882
OY 811 atccatccctctgtctgctcatcggtggtcttccttggttgtaagaccttccaaagcta 870
DB 883 ATCATCTCTCTTGTGTGCTCATCGGGGTCTCTTGTTGTTGAAGCATCTTCCAAAGCTTA 942
OY 871 acgggctgtgaactatctggagcggacccataaacttgagagacactctaactctta 930
DB 943 ACGGCGTGTGAACATATCTGGGACGAGCCATTAACATGAGAGACATCTTAACCTCTTA 1002
OY 931 tctgtcttatctgttactctgttttaagaatgaactaacactatactttctaacagctt 990
DB 1003 TCGTGCTTATCTGTTAACTGTTTAAAGATGACATCACTATATTTTCTAACACCTTT 1062
OY 991 tccaacttaaaccaatcatalaagcttcaattgtctcttggaatgatcgtcttggct 1050
DB 1063 TCCAACTTAACCACTATATAGCTTTTATTTGCTTTTGGAAATATGCTGCTTTTGGT 1122
OY 1051 cggttttagatacttaaacacttaccacactcgtgcacagtttgactcaatgltcaagtaata 1110
DB 1123 CGGTTTAAATCTTAAACACTTTTACCACTGCGCATGTGACATGATGTCAGTCAATA 1182
OY 1111 taacttcaacttgaacaaaatgttatattatattgatattatcttgtaacattaa 1170
DB 1183 TAACTTCAACACTTGAACAAAAGTGTATTTATATGATTATTTGTAACATTAA 1242
OY 1171 gatgttttttccagagctgttctcaatgaagtaactaagatacttaactgttatlaactat 1230
DB 1243 GATGTGTTTTTCCAGCGCTGTTTCATAGTAGTACTAGATATATTAACGTATTAACCTAT 1302
OY 1231 ttccatgatgtcaactgcttaagtttctatgcagaataaattatattta 1283
DB 1303 TTTCCATGATGTCAACTGCTTAAGTTTATATGACGAATAATATATATATTA 1355
RESULT 3
MM55041
LOCUS Mus musculus uracil-DNA glycosylase (ung) mRNA, complete cds.
DEFINITION
ACCESSION U55041
VERSION U55041.1 GI:1762317
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 1857)
Svendsen, P.C., Yee, H.A., Winkfein, R.J. and van de Sande, J.H.
The mouse uracil-DNA glycosylase gene: isolation of cDNA and
genomic clones and mapping ung to mouse chromosome 5
JOURNAL MEDLINE 97311407
MEDLINE 97311407
REFERENCE 2 (bases 1 to 1857)
Svendsen, P.C.
Direct Submission
TITLE University of Calgary, Rm.2153 HSC, 3330 Hospital Dr. N.W.,
JOURNAL Calgary, AB T2N 4N1, Canada
FEATURES
source
1..1857
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="5"
1..27
/gene="ung"
1..1857
/gene="ung"
28..915
/gene="ung"
/function="DNA repair enzyme"
/codon_start=1
/product="uracil-DNA glycosylase"
```

```

/protein_id="AAC53197.1"
/db_xref="GI:1762318"
/translation="MGVLRNRSLRLARRAGLSLNPNDSPDSSRQASPAKKAKEVNEQ
GSPLSAEQVLRQNRKAALRLAARVNPAGGEFWMKQOLCEFPKPFVKLMGVAE
ERNHHKVPPEPOVFTWMDIRVAVVIGDDYVGNQNAHGICFSVORPVPSPS
LENTPEKLSIDGEPVHPGHGDLSGMAQGVLLNAVLTVRHQAQNSHKREMBQPTD
AVSVLNONSLGVFLPLNGSTYQAQKGSVIDRKRHHVLOTAHPSPLSVHGFGLGRHFS
KANELLQKSGKKPINMKEL"
3'UTR
916..1857
/gene="ung"
/polyA_signal
1836..1841
/gene="ung"
BASE COUNT 371 a 495 c 550 g 441 t
ORIGIN

```

```

Query Match      28.4%: Score 364.2; DB 10; Length 1857;
Best Local Similarity 66.8%; Pred. No. 1e-82;
Matches 519; Conservative 0; Mismatches 258; Indels 0; Gaps 0;

```

```

QY 150 gTggaacaaagacgtatcgcacacagcttcagtgagcagctggaagaatgagccaa 209
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 142 GTGAGACCAAGAGCAAGAGCTGCGCGCTCAGAGCGCGAGAGCTGCGCATCCAGAG 201
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 210 aataagaagaagcagcttgacagatlagcaagaagcaacgctgaggttcggagag 269
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 202 AACCAAGCGCGCGCTGCTCAGAGCTCGCGCCGCAACGCTGCTCGCGCTCGCGGAG 261
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 270 acttgagagaagaagctgctgacaggttgaagaagccatctcaacaatgatgctc 329
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 262 AGCTGGAACAGACAGCTGTCGGGGAGTTGCGGAGCCGCTACTCTCTCAAGCTAATGGCA 321
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 330 ttgttagctgatgagagagcgcctcacacgcgtctacccacgcgctgatcaagtgtacgt 389
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 322 TTTGTTGCTGAAGAAGAAAGAACCAACAGTGTATCCGCCGCCGAGCAGGTTCACACA 381
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 390 tcgacagaagatgtgacattcaagaatgtgaagaatgattcttagcgaagacccttc 449
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 382 TGGAGCCCAATGTGCGCATCCGAGATGTGAAGGTGTCTCTGAGACAGATCCCTAT 441
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 450 cagcgtcccaacaaagacatgagctctgttcaagtgtgcaaaagcagattcccccctcc 509
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 442 CACGGACCTAATCAAGCTACAGGGCTCTGCTTCAAGTGTCCAAAGACCAAGTTCACCCCG 501
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 510 cccagctctgtaacatatacaagaatgtgtgtaaccagatgattggtcctaagaacatcct 569
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 502 CCCAGTTTGGAAACATTTTAAAGAGCTGTCTACAGACATCATGTTGTTTCATCT 561
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 570 ggaactgtgagatctaagcgatgaggaacaaagaggtgtcgtcttaacgcgtgtcgt 629
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 562 GGCCTATGGGGATTTGTCAAGGGTGGCGCCGCAAGGTGTCTCTCTCTCAACGCCGCTCTC 621
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 630 accgtgagcggccatcaagcgaactccacaaagacagagagctggaagaccttaccagac 689
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 622 ACTGTCCGGGCCACCACCAAGCAATTCCATTAAGGAGAGGGGCTGGAGCAGTTTCACGGAC 681
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 690 gctgtgataaagtgtgagcggtcaacgcggaagagatcgttctcctgtgtgtgaggtctca 749
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 682 GCGGTGCTGTCTGCTGATGATGAGAACTGAGTGGCTGCTGTCTCTCTCTGCGGCTCT 741
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 750 taagcccaataaagaagagagcaccatcgacaggaagaagctcaccatgtcttgaagctgt 809
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 742 TATGCTCAAGAAAGGCGGTGTCTCATTTGACAGAAAGCCGACCATCTCTCTGACAGACGCT 801
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 810 catcatcctctgtctcatcgttggtgtcctgtgtgtgtgaagacttcccaagagct 869
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 802 CACCCCTCTCCCTGTGCTGTCAGAGAGGGTTCTTGGATGACAGACATTTTCTAAAGCC 861
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 870 aacggagctgtgaaactatcttggaacgagcctataaacttggaagacacttaactac 926
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 862 AATGAACGTGCTCCAGAGATGTGCAAGAGCCCATCACTGGAAGAGCTGTGAGACC 918
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 4
BC011039
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

```

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: amadansystemsbiology.org

Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAX Plate: 18 Row: e Column: 22

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 1650875.

Location/Qualifiers

1..2002

/organism="Mus musculus"

/db_xref="LocusID:22256"

/db_xref="taxon:10090"

/clone="MGC:13929 IMAGE:4009947"

/tissue-type="mammary tumor metastasized to lung.

MMTV-LTR/Mnt1 model. Expression driven by an MMTV-LTR

enhancer."

/clone.lib="NCI-CGAP_Lu30"

/lab.host="DH10B"

/note="Vector: pCMV-Sport6"

126..1046

/codon.start=1

/product="uracil-DNA glycosylase"

/protein_id="AAH1039.1"

/db_xref="GI:15029660"

/translation="MIGOKTLVSFSPPTGKRTTRSPVPYGSVAIEIGDVAASP

AKKARVEONEGSGSPLSAQVLRQNRKAALRLAARVNPAGFSGSMKQOLCEGEGKCF

YFKLMGFAVERHNHKKVYPPPEOVFTWMDIRVAVVILGDDYVGNQNAHGICF

SVORPVPSPSLENTPEKLSIDGEPVHPGHGDLSGMAROGVLLNAVLTVRHQAQNS

HKEREMBOFTDAVSVLNONSLGVFLPLNGSTYQAQKGSVIDRKRHHVLOTAHPSPLSV

HRGFLGCHFSKANELLQKSGKKPINMKEL"

BASE COUNT 417 a 549 c 581 g 455 t

ORIGIN

Query Match 28.3%: Score 362.6; DB 10; Length 2002;

Best Local Similarity 66.7%; Pred. No. 2.7e-82;

Matches 518; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

```

QY 150 gTggaacaaagacgtatcgcacacagcttcagtgagcagctggaagaatgagccaa 209
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 273 GTGAGACCAAGAGCAAGAGCTGCGCGCTCAGAGCGCGAGAGCTGCTCGCATCCAGAG 332
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 210 aataagaagaagcagcttgacagatlagcaagaagcaacgctgaggttcggagag 269
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```


QY	870	aacggcgctgctgaactatctggagcggagcctataaactggagagcacttaactc	926
Db	864	AATGAACCTGCTCCAGAGACTGTGGCAAGAAGCCATCAACTGGAAGAGACTGTGACCC	920
RESULT	6		
LOCUS	MM08975	1951 bp	linear
DEFINITION	M.musculus mRNA for uracil-DNA glycosylase.		
ACCESSION	Y08975		
VERSION	Y08975.1	GI:1850875	
KEYWORDS	ung2 gene; uracil-DNA glycosylase.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1951)		
TITLE	Nilsen, H.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (21-OCT-1996) H. Nilsen, UNIGEN Center for Molecular Biology, University of Trondheim, N-7005 Trondheim, NORWAY		
AUTHORS	2 (bases 1 to 1951)		
TITLE	Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T.A., Skorpen, F., and Krokan, H.E.		
JOURNAL	Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the ung gene		
MEMLINE	Nucleic Acids Res. 25 (4), 750-755 (1997)		
FEATURES	97169285		
SOURCE	Location/Qualifiers		
gene	1..1951		
CDS	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone_lib="Unizap-XR"		
	69..989		
	/gene="ung2"		
	69..989		
	/gene="ung2"		
	/codon_start=1		
	/product="uracil-DNA glycosylase"		
	/protein_id="CA70168.1"		
	/db_xref="GI:1850876"		
	/db_xref="MGD:MG1:109352"		
	/db_xref="SPTREMBL:P97285"		
	/translation="MIGOKTLYSFFSPPTGKRTTRSRBEPYSGVALETGGDAVASB AKKRAVENEQSGSLSAEQLRIORNKAAALRIARVRVPGFESGKQDCEGFGKRYFLVMGVFAERHNNHKKVYPPPEGVFMTCMDLRDKVYVILGGDPYHGMQAGLGE SVQRPVPRPRLNLEIKFELSPDIDGFGYHPCGSDLSGMARQGVILLNVTYRAHANS HKRGMEDQFDVAVSYMLNQLNSGIFLIMGSYIAOKKSIDRKHHVLIQTAAHPSPLSV YRGLGGEHFKSAEELLQKSGKKRNKML"		
BASE COUNT	416 a 523 c 565 g 447 t		
ORIGIN			
Query Match	28.1%; Score 361; DB 10; Length 1951;		
Best Local Similarity	66.5%; Pred. No. 7e-82;		
Matches 517; Conservative	0; Mismatches 260; Indels 0; Gaps 0;		
QY	150	gtggagcaaaaagcgtcctcgccacagctttcagtgagagcagctggaagaaatggccaaa	209
Db	216	GTTGGAGCGAAGACGACCAAGCGCTCGCGGCTCAGCGCCGACGCGGACACTGTGCGCATCCAGAG	275
QY	210	ataaagaagcagcgcttgacaagaattgagcaaaagcaacgcgtgcaagtttcggagag	269
Db	276	AACAAGGCGCGCGCTGCTCAGGCTCCGCGCCGCCAACGTGCTGCGCGCTTCGGCGAG	335
QY	270	acttgagagaagagcgtgctgcagagtttgaagaagacatcttaaaacaattgattgcc	329
Db	336	AGCTGGAAGACGACACTGTGCGGGAGATTTCGGGAAGCCGTAATTGCTCAAGCTTAATGGGA	395
QY	330	tttgagctgtatgaaggagccgttaacacgcgtctaccacccgcgcgtcatcaattgacagt	389

[illegible]

D	384	GGGCGCCCTGTACAGCTGGCGGCCGCACAAGTGGCCCGTGTTGGAGAGAGCTGGAAAG	443
OY	279	agagaagctgctgcacgaattgtgaaaaggcataacttcaaaatgtagtccttgtaagt	338
D	444	AAGCACCCTCACCGGGAGATTGGGAAAACCGATTTTATTCAGACTATGAGTATTCTTGA	503
OY	339	gatbagaggaagccgycacaccgctcaccacagcggtgtcaagtgtaagttgaagttcgacaag	398
D	504	GAGGAAAAAAGACATTTAACACTGTTTTATCCACCCTCCACACCAAGTCTTACCTGGACCCAG	563
OY	399	atgttgtacatccaagaatgltgaaagtagtgaattctcaggccagagaccttacaacaggtccc	458
D	564	ATGTGTGACATAAAGAATGTGAGGTTGTCAATCTCGGGACAGAGATTCATATCATGTGACCT	623
OY	459	aaccaagcacatgtagcttcttgtaagtgtagcaaaaaagccagtlccccctccccagcttc	518
D	624	AATCAAGCTCACGGGCTCTGTTTGTGTTCMAAGCGCTGTTCCCTCCGCTCCGCGAGTTTG	683
OY	519	gtgacacataccaagaattgltacccgaacttgtagtcttcaagaactccttgacaatgta	578
D	684	GAGAACATTTTAAAGAAGTTGTCTACACACATGAGAGATTGTTTCATCTGGCCATGGA	743
OY	579	gacccaagcggatagggcaaacaaaggggtgctcgtcatcaacgsgtgcgtgacgctg	638
D	744	GATTATATCTGGGTGGGCGCAAGCAAGTGTCTCCCTTCACAGCGTCTCCACGGTTCGT	803
OY	639	gcccatcagcccaactcccaacaaggaacagagctgtggagaaccttacaacgagctgtgac	698
D	804	GCCCATCAAGGCCCAACTCTCATAGGAGCGAGGCTGGGAGCAGTTTCATCATGATCAGTTGTG	863
OY	699	aagtgctgagcgcgtacaacgggaaggaatgcgtttccctggtgtgggtcgtcctaagcccat	758
D	864	TCTGGCTTAATTCAGACACTCGAATGGCCTGTGTTTCTGTGCTGGGCTCTTATCTCTCAG	923
OY	759	aagaagggagcagccatcgacagagaacgtcaccaatgcttctcaagctgttcatcatcat	818
D	924	AAGAAGGCGACAGTGCATTTGATGAGAAAGCGGACACATGTACTACAGACGCGTCTATCCCTCC	983
OY	819	ccttctgtctcatcgtgtggttcttctgtgtgttaaagcaacttccaaaggtacaagggctg	878
D	984	CCTTGTCACTGATATAGAGGGTCTTTTGGATGTGAACACTTTTCAAAGCACCAATGAGCTG	1043
OY	879	ctgaaacctatctgggaagcagcctctaaactgtagagcaactcttaa	923
D	1044	CTGCAAGAAGTCTGGCMAAGGCCCATTTGACGTGGAAGAGCTGTGA	1088
RESULT	12		
LOCUS	BC004037	1737 bp	mRNA linear ROD 12-JUL-2001
DEFINITION	Mus musculus, similar to uracil-DNA glycosylase, clone MGC:7655		
ACCESSION	BC004037	IMAGE:3495894, mRNA, complete cds.	
VERSION	BC004037.1	GI:13278470	
KEYWORDS	MGC.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1737)		
JOURNAL	Strausberg, R. Direct Submission Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov		
COMMENT	Contact: MGC help desk Email: cgapbs-ref@mail.nih.gov Tissue Procurement: lothar Henninghausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNU) DNA Sequencing by: Sequencing Group at the Stanford Human Genome		

Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark medepaxi1.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAK Plate: 8 Row: 1 Column: 4.

FEATURES

Source

```

/organism="Mus musculus"
/db_xref="caxon:10090"
/clone="MGC:7655 IMAGE:3495894"
/tissue_type="Mammary tumor WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI-CCAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
192. .791

```

CD5

```

/codon_start=1
/product="Similar to uracil-DNA glycosylase"
/protein_id="AAH04037.1"
/db_xref="GI:13278471"
/translation="MGFAEERNHHKYPPEPEQFTTQMDIDRKYVLLIGDQRPYQ
PNOAGLGFQVRPPRPPLSLENTKSTIDFVPTGMDLDSGMAKGVLLINAVL
TVRAQASHKHEGEQPTDAVWMLNQLSGVFLWGLSGYAAQKGSIVLKKRNHVLQ
TAHSPLELVYRGFGCRHFSKANELLQSGKPINMEL"

```

ORIGIN

397 a 423 c 459 g 458 t

24.9%; Score 319; DB 10; Length 1737;

Best Local Similarity 69.9%; Pred. No. 4.3e-71;
Matches 430; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

Matches 430; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

312 ttcaacaattgattgttcctttgtagctgtagagagccgtcacacacgctaccaccg 371

Db 180 TTAACCAGCTAATGGGATTGTGCTGAGAAAGGAACCAACCAAGGTCATATCCGCC 239

372 actgaatcaatatacaattcaacagaagatgtatgacattcaagatatgaaaagtacgtatt 431

240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

2700

100

100

2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818 2819 2820 2821 2822 2823 2824 2825

532 galggcclccaagcaccggaagaccccaagcggacaaacaaagggcgcg 611

Db 420 GATGGTTTCTCATCCCTGGCCATGGGGATTTCACAGGGTGGGCCCGACCAAGTGTCTC 4/9

612 ctgcttaacgcgtgtgctgacgcgtgcggycccatcagggccaactccacaagacagaggc 671

Db 480 CTCCTCAACGCCGTCTCTACTGTCCGGCCACCACCAAGCCAATTC CATTAAGGAGAGGGGC 539

672 tgggaacccttcaccgacgcctgtatcaadtggcctgaagcgtcaaccgggaagagtcgtt 731

540 TGGGAGCAGCTTCACGGACCGCTGTCCTGATTCAGAACCTGAGTGGCCCTCGTC 599

[illegible]

A vertical ruler with markings from 0 to 10 cm. The markings are in millimeters, with major ticks every centimeter and minor ticks every millimeter. The ruler is oriented vertically with the 0 mark at the bottom and the 10 cm mark at the top.

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

852 aaagcaccclcccaagcclgcgcgaacclalclcggaagcagccclacaadaclg 311

Db 720 AGACATTTTCTTAAGCCATGAGTCTCCAGAGTGTGGCAAGAGCCCATCACTG 779
Qy 912 agagcacttaactc 926
Db 780 AAGGAGCTGTGACCC 794

RESULT 13
CNS01DWM 189672 bp DNA linear PRI 25-MAY-2000
LOCUS Human chromosome 5 DNA sequence *** IN PROGRESS *** BAC R-391B7 of
DEFINITION library RPCR-11 from chromosome 5 of Homo sapiens (Human), complete
sequence.
ACCESSION AL136059.2 GI:7106618
VERSION AL136059.2
KEYWORDS HTG; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 189672)
GENSCOPE.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2000) to the EMBL/GenBank/DBJ databases
COMMENT On Feb 27, 2000 this sequence version replaced gi:6981976.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast vector,
phage, etc. even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the 77 to the 576 end.

Overall quality chart :
Range : bases
0 : 2
1 - 9 : 231
10 - 19 : 880
20 - 29 : 2766
30 - 39 : 7567
40 - 49 : 8585
50 - 59 : 13484
60 - 69 : 28999
70 - 79 : 55767
80 - 89 : 42661
90 - 99 : 28730

Percentage of bases with a quality value >= 40 : 93 %.

FEATURES
source Location/Qualifiers
1..189672
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone.lib="RPCR-11"
/clone.lib="R-391B7"
BASE COUNT 57335 a 40587 c 38988 g 52760 t 2 others
ORIGIN

Query Match 20.5%; Score 262.8; DB 9; Length 189672;
Best Local Similarity 62.8%; Pred. No. 4.3e-56;
Matches 491; Conservative 0; Mismatches 282; Indels 9; Gaps 5;

Qy 169 cgcacagcttccagtgagcagctgaaagaaagcacaagcagcagctg 228
Db 24120 CTCCTCAGCAGCAGCAGCAGCAGTATCTGTCTCCAGAGAAACATGCGCCACAGCCCTGC 24179
Qy 229 acaagattgagcaaaagcagcgtcaggttccgagagacttggagagagagcctg 288
Db 24180 TCAGACTCAGCAGCAGCAGCAGTGTGTC-GTGGGCTTGTGTGAGATTGGAGAGACATCTCA 24238

Qy 289 ctgcagagttgaaagccacttaacaaatgagtccttctgtgagctgagagga 348
Db 24239 GCAGAGATTTGGGAAACCATATTTTAAAGCTCATGATGTGTGCTGAGTAAGGA 24298
Qy 349 gccgtcacaccgtctaccacccggtctgataagtgatcagttgacagagatgtaga 408
Db 24299 AGGATTACACTGTTATCTCTCCACACCAAAATTCACCTGAGCTAGATGTGTACA 24358
Qy 409 ttcagatgtgaaagtagtattcctagagccaccccttaacagc--tcccaacagc 466
Db 24359 TAAGAGATGTGAAGAGATGTCATTCAGAGAGATTCATATGAGACCCCATCAAGC 24418
Qy 467 acatgagcttcttccagtgctgcaaaagccaggttccctcccccagctctgtgaacat 526
Db 24419 TCGTGGGCTGCTTTAGTCTTCAAAAGCCTGTTCACCTCTGCCAGTTGGAAAAAT 24478
Qy 527 atcaagaatgtgtacacattgagcttcaagcctcagagatgagatcgaag 586
Db 24479 TTATTAAGACTGTCTACACATAGATGCTTTGTTCAATCCAGCCATGAGATGTATC 24538
Qy 587 cggatggcacaacaaaggggtgctgcttaacgcggtgctgacgctgcgggcccata 646
Db 24539 TGGGTGGGTCAACCAAGGTG--TTTACTCAACGCTGTCTCAGTCCAGCCATCA 24595
Qy 647 ggcacactcccaagaagagcagagctgagag--acctcacagcgtgtgacagtg 704
Db 24596 AGCCAAATTCATGAGAGAGAGAGCTGGAGCAATCTCTTAATGCACTGTGCTCTG 24655
Qy 705 ctgagctcaacgcggagagagagctgttctctgttggggctcatagccataagaag 764
Db 24656 CTAAAGCCAGAACCTCATGAGCTGTCTTCTGCTGCGGGCTTCTTATGCTCAGAGAGG 24715
Qy 765 gtagcgacacatcgacagaaagcgtacacatgcttgaagctgttcatcatcctctt 824
Db 24716 ACAGTG-CCATGAGAGAGAGAGCGGACCATGATCTACTGCAAAAGTCTCTCTGTTG 24774
Qy 825 tctgtcatgtgtgttcttctgttgaagcacttcccaagcttaacgagctgagctgaa 884
Db 24775 TCAGCGTATGATGGTTCTTGTGGATGTAGACACTTTCTTAAGCAATGAGCTGCTCAG 24834
Qy 885 ctatctggagcggagcctataaactgagagacacttaactcttattgctccttact 944
Db 24835 AAGTCAGCAGCAAGAAAGCTTACTGACCGGAGAGAGAGTGTGATCTCAATGAGAGTACTCT 24894
Qy 945 gt 946
Db 24895 GT 24896

RESULT 14
AC010258 213353 bp DNA linear HTG 20-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTC-452117, WORKING DRAFT SEQUENCE,
DEFINITION 17 unordered pieces.
ACCESSION AC010258
VERSION AC010258.3 GI:13699482
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 213353)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
AUTHORS
REFERENCE 2 (bases 1 to 213353)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 20, 2001 this sequence version replaced gi:9964744.
-----Genome Center
Center: Joint Genome Institute

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2002, 03:54:35 ; Search time 270.17 Seconds

(without alignments)
8153.389 Million cell updates/sec

Title: US-09-758-017A-1

Perfect score: 1283

Sequence: 1 gacatccgctgcaaatatg.....agaataataatatttta 1283

Scoring table: IDENTITY_NUC

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : _N_Geneseq_032802:*

1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1278.2	99.6	1283	22	AAS09498
2	1164.2	90.7	1355	22	AAS09499
3	335.4	27.7	1967	22	AAH34344
4	335.4	27.7	2053	18	AAT73564
5	335.4	27.7	2122	24	AAS94978
6	207.6	16.2	1532	22	AAS63241
7	180	14.0	657	22	AAF94351
8	172.2	13.4	1830121	17	AAT42063
9	172.2	13.4	349980	21	AAF21609

10	172.2	13.4	349980	21	AAF21610
11	160.8	12.5	2469	20	AAH13540
12	159.6	12.4	681	20	AAH27774
13	147.8	11.5	8753	12	AAH32851
14	142.2	11.1	1348	21	AAH36561
15	138	10.8	2944528	24	ABA03041
16	137.4	10.7	11340	19	AAV52280
17	132.2	10.3	2988	23	AAH75904
18	121.4	9.5	17846	18	AAV74420
19	111.4	8.7	2975	22	AAH53981
20	111.4	8.7	4114	22	AAH54820
21	111.2	8.7	651	22	AAH52516
22	107.4	8.4	133719	21	AAH64754
23	107	8.3	335100	20	AAV78803
24	107	8.3	137507	19	AAV19941
25	106.4	8.3	487	21	AAH78326
26	102	8.0	4100	20	AAH32019
27	102	8.0	4100	22	AAH90076
28	100	7.8	119211	22	AAH28553
29	97.8	7.6	910715	20	AAH20248
30	94	7.3	1230025	20	AAH91990
31	89	6.9	4403765	22	AAH96683
32	89	6.9	4411529	22	AAH96682
33	83	6.5	7361	19	AAV62153
34	83	6.5	8852	19	AAV62130
35	83	6.5	117213	19	AAV62176
36	83	6.5	154746	24	AAH25519
37	80	6.2	500	22	AAH41150
38	78.2	6.1	4881	19	AAV63438
39	68	5.3	124884	22	AAH74201
40	68	5.3	125157	22	AAH74202
41	67.4	5.3	777	19	AAH14447
42	61.6	4.8	13228	23	AAH55967
43	49.4	3.9	762	22	AAH66426
44	49.4	3.9	349980	22	AAH68528
45	48.4	3.8	1262	23	AAH65031

ALIGNMENTS

RESULT 1	
AAS09498	
ID	AAS09498 standard; CDNA; 1283 BP.
AC	AAH34344;
XX	
DT	24-OCT-2001 (first entry)
XX	
DE	Atlantic cod cDNA encoding heat-labile uracil-DNA glycosylase, UNG #1.
XX	
KW	Atlantic cod; heat-labile uracil-DNA glycosylase; UNG; UNG.
KW	PCR control; LCR control; ligase chain reaction; carry-over prevention.
XX	
OS	Gadus morhua.
XX	
XX	
FT	Key
FT	Location/Qualifiers
FT	18..923
FT	/*tag= a
FT	/Product= "UNG #1"
XX	
PN	WO200151623-A1.
XX	
PD	19-JUL-2001.
XX	
PF	10-JAN-2001; 2001WO-NO00008.
XX	
PR	12-JAN-2000; 2000NO-0000163.
XX	
PR	27-OCT-2000; 2000NO-0005428.
XX	
PA	(BIOT-) BIOTEC ASA.
XX	

Neisseria meningit
Enterococcus faec
B. pallidus uracil
P. denitrificans Ec
Arabidopsis thalia
Listeria monocytog
Streptococcus pneu
DNA encoding novel
Staphylococcus aur
S. epidermidis gen
S. epidermidis ope
Macaca mulatta rha
KSHV LUR DNA (nucl
KSHV long unique c
Human cancer assoc
Human MERT1 relate
L1385 cDNA clone.
Genomic fragment #
Borrelia burgdorfe
Nucleotide sequenc
Mycobacterium tube
Mycobacterium tube
HSV-2 strain SB5 C
HSV-2 strain SB5 C
HSV-2 strain SB5 C
Human herpesvirus
Left nucleic acid
Feline herpesvirus
Nucleotide sequenc
H. pylori GHP0 902
Propionibacterium
C glutamicum codin
C glutamicum codin
DNA encoding novel

PI Lanes O, Willasen NP, Guddal PH, Gjellervik DR;
 XX WPI: 2001-451854/48.
 DR P-PSDB: AAU04939.
 XX
 PT New cod liver uracil-DNA glycosylase enzyme, useful in monitoring or
 PT controlling a reaction system multiplying DNA sequences or in
 PT carry-over prevention procedures
 PS
 XX Claim 7: Page 52-54; 59pp; English.
 CC The sequence encodes an Atlantic cod heat-labile uracil-DNA glycosylase,
 CC (UNG/UDG). The enzyme is useful in monitoring and/or controlling a
 CC reaction system multiplying DNA sequences, e.g. PCR (polymerase chain
 CC reaction) or LCR (ligase chain reaction). The enzyme is also useful in
 CC carry-over prevention procedures.
 XX
 SQ Sequence 1283 BP; 364 A; 278 C; 280 G; 360 T; 1 other;

Query Match 99.6%; Score 1278.2; DB 22; Length 1283;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1280; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gacatccggttgaataatggttgaattggtggtatgacccaagaatcatatcata 60
 Db 1 gacatccggttgaataatggttgaattggtggtatgacccaagaatcatatcata 60
 QY 61 atcgggtgttacaggtttactaatccccaacttatctatgttttctaataataaga 120
 Db 61 atcgggtgttacaggtttactaatccccaacttatctatgttttctaataataaga 120
 QY 121 taacgcgaagaacttgaggtccctcaaatgttgaacaaaagcgtatcgccaagctt 180
 Db 121 taacgcgaagaacttgaggtccctcaaatgttgaacaaaagcgtatcgccaagctt 180
 QY 181 caggtgagagagcttggaataatggccaataataagaagcagcgttgagaattagag 240
 Db 181 caggtgagagagcttggaataatggccaataataagaagcagcgttgagaattagag 240
 QY 241 caaagcaacgcttcaggttttcgagagagacttggaagaagctgctgcagagttg 300
 Db 241 caaagcaacgcttcaggttttcgagagagacttggaagaagctgctgcagagttg 300
 QY 301 aaaagccatacttcaaaacttgatgtccttctgtatgataagagagcgttcacacg 360
 Db 301 aaaagccatacttcaaaacttgatgtccttctgtatgataagagagcgttcacacg 360
 QY 361 tctaccacacgctgatacaagtgtacagttgacagaagaatgtgtgacattcaagatgta 420
 Db 361 tctaccacacgctgatacaagtgtacagttgacagaagaatgtgtgacattcaagatgta 420
 QY 421 aagtagtgattctagccgagaccccttaccacggtccccaagaacatgatacttgtt 480
 Db 421 aagtagtgattctagccgagaccccttaccacggtccccaagaacatgatacttgtt 480
 QY 481 tcaagtgtgcaaaagccagttcccccctcccccagctctgtgaacataataaagaattgt 540
 Db 481 tcaagtgtgcaaaagccagttcccccctcccccagctctgtgaacataataaagaattgt 540
 QY 541 gtaacgacattgattggtcttcaagcatcctgagacatggaatttaagcgaatggcaaac 600
 Db 541 gtaacgacattgattggtcttcaagcatcctgagacatggaatttaagcgaatggcaaac 600
 QY 601 aaggggtgtgtgtcttaacgagcgtgtacacgttgcggccatcagccaactcccaaca 660
 Db 601 aaggggtgtgtgtcttaacgagcgtgtgtacacgttgcggccatcagccaactcccaaca 660
 QY 661 aggaacagagctgtgagacattcaccagacgtgtgtatcaagtgtgacgttcaaccggg 720
 Db 661 aggaacagagctgtgagacattcaccagacgtgtgtatcaagtgtgacgttcaaccggg 720
 QY 721 aagagtgcttttctctgtgtgtggtcctacacgcccataagaaggagcgaccatcgcaca 780

Db 721 aagagtgctttccctgtgtgtggtcctacacgcccataagaaggagcgaccatcgcaca 780
 QY 781 ggaacgctacacatgtctgcaagcgtgttcacatccatccttctgtgtcgtacgtgggt 840
 Db 781 ggaacgctacacatgtctgcaagcgtgttcacatccatccttctgtgtcgtacgtgggt 840
 QY 841 tccctgtgtgaagcaacttccccaagcgtgacggtgtgtgaactatctggagcggagc 900
 Db 841 tccctgtgtgaagcaacttccccaagcgtgacggtgtgtgaactatctggagcggagc 900
 QY 901 ctataacgaggagacacttaacttctatgtctgcttactactctgttaactgtttaaga 960
 Db 901 ctataacgaggagacacttaacttctatgtctgcttactactctgttaactgtttaaga 960
 QY 961 tgaacatcacatatactttctacagcgttttccaagttccaagttcaaaccaactataagcttca 1020
 Db 961 tgaacatcacatatactttctacagcgttttccaagttccaagttcaaaccaactataagcttca 1020
 QY 1021 ttgtcttttgaagaatgagctgtcttggctggttttagatatacttaaacacttaaccac 1080
 Db 1021 ttgtcttttgaagaatgagctgtcttggctggttttagatatacttaaacacttaaccac 1080
 QY 1081 tctgcacatgtgacatgtttcagtcataataacttcaacacttgaacaaaatgttat 1140
 Db 1081 tctgcacatgtgacatgtttcagtcataataacttcaacacttgaacaaaatgttat 1140
 QY 1141 ttataatgtatatacttctgacatlaaagaatgttttttcccaagcgtgttcaagg 1200
 Db 1141 ttataatgtatatacttctgacatlaaagaatgttttttcccaagcgtgttcaagg 1200
 QY 1201 tactagatataataactgttatataactatttccatgtgtgacatgttgaatttta 1260
 Db 1201 tactagatataataactgttatataactatttccatgtgtgacatgttgaatttta 1260
 QY 1261 tgcagaataaattatattta 1283
 Db 1261 tgcagaataaattatattta 1283

RESULT 2
 ID AAS09499 standard; cDNA; 1355 BP.
 XX
 AC AAS09499;
 XX
 DT 24-OCT-2001 (first entry)
 DE
 XX Atlantic cod cDNA encoding heat-labile uracil-DNA glycosylase, UNG #2.
 KW Atlantic cod; heat-labile uracil-DNA glycosylase; UNG; UDG;
 KW PCR control; LCR control; ligase chain reaction; carry-over prevention;
 SS.
 OS
 XX Gadus morhua.
 FH
 FT Key Location/Qualifiers
 FT CDS 90..995
 FT /*tag= a
 FT /product= "UNG #2"
 XX
 PN WO200151623-A1.
 XX
 PD 19-JUL-2001.
 XX
 PF 10-JAN-2001; 2001WO-NO00008.
 XX
 PR 12-JAN-2000; 2000NO-0000163.
 PR 27-OCT-2000; 2000NO-0005428.
 XX
 PA (BIOT-) BIOTEC ASA.
 XX
 PI Lanes O, Willasen NP, Guddal PH, Gjellervik DR;

XX WPI: 2001-451854/48.
DR P-PSDB: AAU04940.
PT New cod liver uracil-DNA glycosylase enzyme, useful in monitoring or
PT controlling a reaction system multiplying DNA sequences or in
PT carry-over prevention procedures
XX
XX
PS Claim 7, Page 54-56; 59pp; English.
XX
CC The sequence encodes an Atlantic cod heat-labile uracil-DNA glycosylase,
CC (UNK/UNG). The enzyme is useful in monitoring and/or controlling a
CC reaction system multiplying DNA sequences, e.g. PCR (polymerase chain
CC reaction) or LCR (ligase chain reaction). The enzyme is also useful in
CC carry-over prevention procedures.
XX
SQ Sequence 1355 BP; 393 A; 287 C; 304 G; 371 T; 0 other;

Query Match 90.7%; Score 1164.2; DB 22; Length 1355;
Best Local Similarity 98.5%; Pred. No. 0;

Matches 1175; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 91 aaacttaagtcttcttaattgaataacgcgagagaactgagctcctaactg 150
DB 163 aaaccgaaaagcatgcccgaagaagltcagataacgcgagaactgagctcctaactg 222
QY 151 tggaaacaaagacgctcaccagcttcagtgagagcagctggaaagaatggccaaa 210
DB 223 tggaaacaaagacgctcaccagcttcagtgagagcagctggaaagaatggccaaa 282
QY 211 ataagaaagcagcgttgaagaagattgaagcaaaagcagcgttcgagagaga 270
DB 283 ataagaaagcagcgttgaagaagattgaagcaaaagcagcgttcgagagaga 342
QY 271 ctggagaagaagcgtgctgcagaggttgaaaagccatactccaacatgattcct 330
DB 343 ctggagaagaagcgtgctgcagaggttgaaaagccatactccaacatgattcct 402
QY 331 ttgtatgctgataagagagcggtcacaccgtctaccaccggtgataagtgatacgtt 390
DB 403 ttgtatgctgataagagagcggtcacaccgtctaccaccggtgataagtgatacgtt 462
QY 391 cgacagagatggtgacatctcaagaatgtgaagtagtgaattgaagccagagccctacc 450
DB 463 ggaacagagatggtgacatctcaagaatgtgaagtagtgaattgaagccagagccctacc 522
QY 451 acggtcccaacccaagacatgagctctgttcagtggtgcaaaagccaagtcctccctccc 510
DB 523 acggtcccaacccaagacatgagctctgttcagtggtgcaaaagccaagtcctccctccc 582
QY 511 ccaagtcctgatacatatacaagaattgtgtacagacattgtggtctcaagcatctcg 570
DB 583 ccaagtcctgatacatatacaagaattgtgtacagacattgtggtctcaagcatctcg 642
QY 571 gacatggagatcgaagcgaatggaacaaaggggtgctgtgcttaacgagcggtcga 630
DB 643 gacatggagatcgaagcgaatggaacaaaggggtgctgtgcttaacgagcggtcga 702
QY 631 ccgtgagggcccatcagacagcaactcccaacaaagagagcgtggagagactcaccagag 690
DB 703 ccgtgagggcccatcagacagcaactcccaacaaagagagcgtggagagactcaccagag 762
QY 691 ctgtatataagtggtgagcggtcaacccgggaagggatcttctcgtgttggtggtctat 750
DB 763 ctgtatataagtggtgagcggtcaacccgggaagggatcttctcgtgttggtggtctat 822
QY 751 agccccaataagaagagagacatcgaacaggaagaacgtaccatgcttgcgaagctgttc 810
DB 823 agccccaataagaagagagacatcgaacaggaagaacgtaccatgcttgcgaagctgttc 882
QY 811 atccatctccttctgtcgtcactcgtgggtctcctgtgtgttaagaacatctccaagacta 870

DB 883 atccatctccttctgtcgtcactcgtgggtctcctgtgtgttaagaacatctccaagacta 942
QY 871 acgggtcgtcgaactatacttggagcggagccataaactcggagagcactcctaactt 930
DB 943 acgggtcgtcgaactatacttggagcggagccataaactcggagagcactcctaactt 1002
QY 931 tctgtccttaactggttaactggttgaagatgaacacacatacttcttaacagctt 990
DB 1003 tctgtccttaactggttaactggttgaagatgaacacacatacttcttaacagctt 1062
QY 991 tccaagtccaacccaactcataaagcttcatgtcttcttggatgatgctgcttgg 1050
DB 1063 tccaagtccaacccaactcataaagcttcatgtcttcttggatgatgctgcttgg 1122
QY 1051 cgggtttagatacttaaaacacttaccactcgtgcacatgttgatacgttcaagtaata 1110
DB 1123 cgggtttagatacttaaaacacttaccactcgtgcacatgttgatacgttcaagtaata 1182
QY 1111 taacttcaacttgaacaaaatgttatattatattatattatattatattatattat 1170
DB 1183 taacttcaacttgaacaaaatgttatattatattatattatattatattatattat 1242
QY 1171 gatgttttttccagagcgtttcactaagtaagtaagataactgttataactt 1230
DB 1243 gatgttttttccagagcgtttcactaagtaagtaagataactgttataactt 1302
QY 1231 ttccatgatgtcaactgcttaagttttatgcagaataaattatattat 1283
DB 1303 ttccatgatgtcaactgcttaagttttatgcagaataaattatattat 1355

RESULT 3
AAH34344
ID AAH34344 standard; CDNA; 1967 BP.
XX
XX
AC AAH34344;
XX
XX
DT 03-SEP-2001 (first entry)
XX
XX
DE Human colon cancer antigen encoding CDNA SEQ ID NO:1426.
XX
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 12; ss.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200122920-A2.
XX
XX
PD 05-APR-2001.
XX
XX
PE 28-SEP-2000; 2000MO-US26524.
XX
XX
PR 29-SEP-1999; 990US-0157137.
XX
XX
PR 03-NOV-1999; 990US-0163280.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX
DR WPI: 2001-235357/24.
XX
XX
DR P-PSDB: AAG74939.
XX
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX
PS Claim 1; Page 3119; 9803pp; English.
XX
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytosolic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication. meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 1967 BP; 504 A; 438 C; 500 G; 524 T; 1 other;

Query Match 27.7%; Score 355.4; DB 22; Length 1967;
 Best Local Similarity 66.5%; Pred. NO. 1.4e-88;
 Matches 509; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

QY 159 aagacgtacatcgccacagcttcagtgagcaagtgaaagaaatgagcaaaataagaa 218
 DB 116 acgcgcctcctcctcgcgcgtgagtcgcgcagcttgaccgcagtgaaagaaagcc 175
 QY 219 gcagcgttcgacaaagattagacaaagcaagcctgcaggttcggagagacttgaga 278
 DB 176 gcgcgcctcctcagactcgcgcgcgcgaacgtgcgcgttgagtgagagagactgaga 235
 QY 279 agagagctgctcagagattgaaagcaatctcaaaatgtgtctcttgatgc 338
 DB 236 aagaccctcagcggagagtcgggaaacgcgtattctcaagctaatggattgttgca 295
 QY 339 gatgagaggagccgtcacccgcgtctacccacgcgcgtatcaatgtacagtgcaagag 398
 DB 296 gaagaaagaaagattacactgtttacaccccaacaaatctcacttgagaccac 355
 QY 399 atgtgtgacattcacaagtgtgaagtgtatgtatctagggcagacccttccagagtc 458
 DB 356 atgtgtgacataaagatgtgaagtgatctcctcgtgagacatcatatcatgacact 415
 QY 459 aaccagacatgactcgttctcagtgatgtgcaaaagccagttccctcccccagctc 518
 DB 416 aacaaagctcagcgcctcgtctttagtctcaaaagccgttcgcctccgcagcttg 475
 QY 519 gtgaacatacaaaagattgtgtacccagatgagcttcaagacatccctgacatgga 578
 DB 476 ggaacattataaagagtggtctacagacatagagatttttcatctcgtgcacgga 535
 QY 579 gatctaagcggatgggcaaaagaggtgtgtgtcttaacgcggtgtgacgcgtgcgg 638
 DB 536 gatttatcgggtggtgcaagcagagtggtctcctctcctaagcgtctccacagcttgc 595
 QY 639 gccatcaagcccaactcccaagaagcagagctgtggagaccttccacagcgtctgac 698
 DB 596 gccatcaagcccaactcccaagaagcagagctgtggagaccttccacagcgtctgac 655
 QY 699 aagtgctagcgtcacaacgggaagagtcgttctccttgggggagctacagcccat 758
 DB 656 tctgtctaataatcagaacatcgaaatggcctgtttcttctgtcctcggggtccttctg 715
 QY 759 aagaagggagcgacacatcgacagaaagtcacacatgtcttgcagcttgcacatc 818
 DB 716 aagaagggagcgacatcgatagaaagcgcacacatgactactacagcgcgtcctcc 775
 QY 819 ccttctgcctcatcgtgtgttctccttctgttgaacacttcccaagagtaagcgctg 878
 DB 776 ccttctgcctcatcgtgtgttctccttctgttgaacacttcccaagagtaagcgctg 835
 QY 879 ctgaaactcttgggagcgagccttaactgagagcactctaa 923
 DB 836 ctgcaagaagtctggcagaagaccatgactggaagagactgtga 880

RESULT 4
 AAT73564
 ID AAT73564 standard; cDNA; 2053 BP.
 XX
 AC AAT73564;
 XX
 XX 28-SEP-1997 (first entry)
 DT
 XX Uracil DNA glycosylase UNG2 cDNA.
 DE
 XX Uracil DNA glycosylase; UNG2; mutagenesis; DNA sequencing;
 KW DNA modification; cell killing; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS /tag= a
 FT mRNA 71..202
 FT /tag= b
 FT /note= "Claim 16"
 FT mRNA 121..130
 FT /tag= c
 FT /note= "Claim 1"
 PN WO9725416-A2.
 XX
 PD 17-JUL-1997.
 XX
 PE 09-JAN-1997; 97MO-GB00057.
 XX
 PR 09-JAN-1996; 96GB-0000384.
 XX
 PA (DZIE/) DZIEGLEMSKA H E.
 PA (NYFO-) NYFOTER AS.
 XX
 PI Kavli B, Krokan HE, Mol CD, Slupphaug G, Tainer JA;
 PT WPI: 1997-372857/34.
 DR P-PDB: AAM21814.
 XX
 PT DNA glycosylase capable of releasing cytosine, thymine or uracil
 bases from DNA - useful in vitro and/or vivo mutagenesis systems
 to remove contaminating DNA prior to PCR amplification
 XX
 PS Claim 1: Page 47-48; 60pp; English.
 XX
 CC A cDNA clone (AAT73564) codes for a new form (UNG2) (AAM21814) of
 uracil DNA glycosylase that is capable of releasing uracil bases
 from single and/or double stranded DNA. It was isolated from a
 human NT2 neuronal precursor cell cDNA library and represents an
 alternatively spliced form of the UNG gene (see also AAT73567) that
 arises by splicing of a previously unrecognized exon (exon 1A) into
 a consensus splice site after codon 35 in exon 1B (previously
 designated exon 1). UNG2 polypeptide differs from UNG1 in the
 presence of a 44-amino acid N-terminal presequence (see also
 AAM21817). Mutagenesis of UNG2 cDNA provides nucleic acids that
 encode cytosine and thymine DNA glycosylases (see also AAM21815-16).
 CC Recombinant DNA glycosylases can be expressed in host cells for use
 in mutagenesis, to remove contaminating DNA prior to PCR, in DNA
 modification and in cell killing.
 CC
 XX
 SO Sequence 2053 BP; 488 A; 486 C; 538 G; 541 T; 0 other;

Query Match 27.7%; Score 355.4; DB 18; Length 2053;
 Best Local Similarity 66.5%; Pred. NO. 1.4e-88;
 Matches 509; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

QY 159 aagacgtacatcgccacagcttcagtgagcaagtgaaagaaatgagcaaaataagaa 218
 DB 248 acgcgcctcctcctcgcgcgtgagtcgcgcagcttgaccgcagtgaaagaaagcc 307

QY 219 gcaagcgttacaagattagagcaaaagcaagccgtgcaagtttcggaagacttgaga 278
DB 308 ggcgcctctgctagactcgcgcgcgaagctgcccgtggttggaagactggaag 367
QY 279 agagagctgctcagagatttgaagccacttcaaaatgtatgtcttctgtagct 338
DB 368 aagcaactcagcgggaggttcgggaacccgtatcttatacaagatggatctgtgca 427
QY 339 gatgagaggagccgtcacaccgtctacccacccgctgataagtgatacttcagagag 398
DB 428 gaagaaagaaagacttactactgtttatccacccacacaaagcttctcacttgaccag 487
QY 399 atgtgtgacattcaagatgtgaaagtatgtattcttagaccagacccttaccagttccc 458
DB 488 atgtgtgacataaaagatgtgaaagtgttcatccttggaagagatcatatcatgagact 547
QY 459 aaccaagcacatgactctcttccagtgtgcaaaagccagttccctcccccagctc 518
DB 548 aatcaagctcaagggcctgctttagtcttaagaaagccctgtctccgcctccagtttg 607
QY 519 gtgaacatatacaaaagatgtgttaccagacatgtatgcttcaagacatcccttgacatgga 578
DB 608 gagaacattataaagatgtgtcttaacagacatagagatttcttcatccttgccatgga 667
QY 579 gatctaagcgtatgagcaaaagagggtgtcgtcgtcttaacgcggtgtcgaacgtcgga 638
DB 668 gatctatctgtgtggtgccaagaggtgtctctcttccatcgaacgtgctccacaggttcgt 727
QY 639 gccacacagggcaaccccccacagagacagagcgtggaaccccttccagacgtgtgac 698
DB 728 gccacacagggcaaccccccac 787
QY 699 aatggtcgtgacgclcaacgcgggaagagtcgttctcgtgtgtgtgtgtgtgtgtgtgtgt 758
DB 788 tccgtgtctaatacagaactcgaatgtgctgttctgtctgtgtgtgtgtgtgtgtgtgtgtgt 847
QY 759 aagaagggagagcagcctcgcagcaggaagacgtcagacatgtcttgaagtgcttccatctc 818
DB 848 aagaagggagcagcctcgcagcaggaagacgtcagacatgtcttgaagtgcttccatctc 907
QY 819 ccttctgtcgtacatcgt 878
DB 908 ccttctgtcgtacatcgt 967
QY 879 ctgaaactatctgtgagcagagcctataaactgtgagagactctaa 923
DB 968 ctgcaagaagctgtgcaagaagcccatgtgactgtgaaagagctgtga 1012

RESULT 5
AAS94978
ID AAS94978 standard; DNA; 2122 BP.
XX
AC AAS94978;
XX
DT 14-FEB-2002 (first entry)
XX
DE Human DNA sequence #233 expressed during foam cell differentiation.
XX
KW Human: foam cell differentiation; atherosclerosis; cerebral stroke;
XX
KW cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200177389-A2.
XX
PD 18-OCT-2001.
XX
FE 04-APR-2001; 2001WO-US11128.
XX
PR 05-APR-2000; 2000US-195106P.
XX

PA (INCY-) INCYTE GENOMICS INC.
XX Shiffman D, Somogyi R, Lawn R, Selthamer JJ, Porter GJ, Mikita T;
PI Tai J;
PI
DR WPI; 2002-010925/01.
XX
XX
PT Composition useful for diagnosis of conditions, disorders or diseases
PT associated with atherosclerosis, comprises several polynucleotides that
PT are differentially expressed in foam cell development -
PS
PS Claim 1, Page 277; 315pp; English.
XX
XX The present invention relates to the isolation of human polynucleotide
CC sequences that are differentially expressed during foam cell
CC differentiation. The polynucleotide sequences of the invention or a
CC composition comprising these polynucleotides are useful as a high
CC throughput method for detecting altered expression of one or more
CC polynucleotides in a sample. The polynucleotides can be used in the
CC diagnosis of disorders associated with foam cell development such as
CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as
CC coronary artery disease. The polynucleotide sequences can also be used
CC as PCR primers and probes. The polynucleotides of the invention are also
CC useful in gene therapy. AAS94746-AAS95021 represent the human
CC polynucleotide sequences of the invention which are differentially
CC expressed during foam cell differentiation.
SQ
SQ Sequence 2122 BP; 487 A; 504 C; 563 G; 568 T; 0 other;
Query Match 27.7%; Score 355.4; DB 24; Length 2122;
Best Local Similarity 66.5%; Pred. NO. 1.4e-88;
Matches 509; Conservative 0; Mismatches 256; Indels 0; Gaps 0;
QY 159 aagacgtcatccgcacacagcttccagctgtgagcagctggaagaaatgccaataaagaa 218
DB 324 acgcgcctccctcgcgcctcgc 383
QY 219 gcaagcgttacaagatttaagcaaaagcaagcgtcgaagtttcggaagacttgaga 278
DB 384 ggcgcctctgtcagactcgc 443
QY 279 agagagctgctgcaagatttgaagacacacttcaaacatgtatgtcttctgtgct 338
DB 444 aagcactcagcgggaggttcgggaacgcgtatcttcaagcctaagtgatttgtgca 503
QY 339 gatgagaggagcgcgtcaccagcttaccacgcgctgatacaagtgtacatgacagag 398
DB 504 gaagaaagaaagacttactactcgtttatccacccacacacacacacacacacacacacac 563
QY 399 atgtgtgacattcaagatgtgaaagtatgtattcttaggccaagacccttaccaggtccc 458
DB 564 atgtgtgacataaaagatgtgaaagtgtgtatcctctgggaacagatcatatcatgagact 623
QY 459 aaccaagcacatgactctgttccagtgtgcaaaagccagttccctcccccagctc 518
DB 624 aatcaagctcaagggcctcgtctttagtctcaaaagccgtctccgcctccgcccagtttg 683
QY 519 gtgaacatatacaaaagatgtgtaccagacatgtatgcttcaagacatcccttggaacgga 578
DB 684 gagaacattataaaggtgtgtcaccagacatagagatttcttctcctcgtgcaatgga 743
QY 579 gatctaagcgtatgagcaaaagagggtgtcgtcgttcaacgcggtgtcgaacgtcgga 638
DB 744 gatttactgtgtgtgccaagaggtgtctccttccatccacagctgtcccaaggttctgc 803
QY 639 gcccatcagcgaacttcccaaggaagagagcgtgggaagccttaccagcgtgtgtatc 698
DB 804 gcccatcagcgaacttcccaaggaagagagcgtgggaagccttaccagcgtgtgtatc 863
QY 699 aatgtgtgagcgtcaacgcgggaagagagtcgttctcgtgtgtgtgtgtgtgtgtgtgtgtgt 758
DB 864 tccgtgttaatacagaactcgtgaatgacctgttctcgtcgtcgtgtgtgtgtgtgtgtgtgtgt 923


```

XX 15-FEB-2001.
XX
XX 03-AUG-2000; 2000MO-US21176.
XX
XX 04-AUG-1999; 99US-0368382.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Chovan LE, Hessler PE, Relch KA;
XX
XX WPI: 2001-147511/15.
XX
XX P-PSDB; AAB88498.
XX
XX Essential bacterial genes from Haemophilus influenzae and methods for
XX identifying 'essential' genes that may be potential therapeutic targets
XX
XX
XX PS Disclosure: Page 52-53; 185pp; English.
XX
XX AAP94345 to AAP94409 represent essential bacterial genes from
XX Haemophilus influenzae, which encode the proteins given in AAB88492 to
XX AAB88556. The present invention also describes methods for identifying
XX essential bacterial genes (i.e. those essential to the survival of a
XX bacterium) using a transposition system. The methods are used to
XX identify essential genes from bacteria, especially H. influenzae (which
XX causes otitis media, meningitis and upper respiratory tract infections)
XX which may be used as targets for potential antimicrobial agents.
XX CC AAP94410 to AAP94416 represent PCR primers used in the exemplification
XX of the present invention.
XX
XX SO Sequence 657 BP; 213 A; 138 C; 130 G; 176 T; 0 other;

Query Match          14.0%; Score 180; DB 22; Length 657;
Best Local Similarity 57.4%; Pred. No. 6.1e-40;
Matches 324; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

OY 348 agccgtcacacgcgtctaccacgcgctgataagtgatgacagagatgtgtgac 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88 agcgggaaacgattatcccccacagaagatgtatttaacgattcaataatactgct 147

OY 408 atcaagaatgtgaagaagatgattcttagcgacagacccttaacgcgttcccaacga 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 148 ttgagagatgtaaagtgtaattttagtgacagatcccttaacgacaacacagcg 207

OY 468 catgactcgtttcaagtgtgcaaaagccagttccctcccccagctcgtgaataa 527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 208 caagcgttgcttttcaagtaaaacgtgaagtagcattcccttccattataata 267

OY 528 tacaagaatgtgtacgacatgtatgcttcaagatcccttgacatgtgaataagc 587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 268 tataaagaatcacacacagatatttcggtatttcaaatgcatcaaatgtattatg 327

OY 588 ggtgtggcaaaaaggggtgctgtcttaacgcggtgtgacagcggtggccacag 647
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 328 aaatggcgcaaaaaggggtatgtctacttaacgtgtgttacttggaagcgatg 387

OY 648 gcaaacctcccaagaagcgatgtagaagaccttccacgacgctgtgataagtgctg 707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 388 gcaattcacacgcgaatttaagtgtagaaggtttacagataagttatgtgcgtactc 447

OY 708 agcgtcaacgcgggaagagtcgtttccctgtgtgggctcattacgcccataagaagga 767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 448 aatgaacatcgtgaaaacgtgttcttacttggggcagtcacgacaacaaaagg 507

OY 768 ggcagcctcgacaggaagatccacatgctcttgcaagctgttcatcattccttctt 827
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 508 caaatgtatgcacgcatcgtcacttgtttaaagctccgcacgtccttcccgltgca 567

OY 828 gctcatgttggttctgtgtgtgtgtaagcacttcccaagcgagcggtcgtcaacta 887
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 568 gcaacatgaaagtttcttctgtgtgtgtcgtcattttcccaaaaacattcatatttggaaagc 627

```

```

OY 888 tctggacggagcctataactgg 911
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 628 cacggaataaacgcgatactg 651

RESULT      8
AAT42063
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
XX OS Haemophilus influenzae.
XX
XX PN WO9633276-A1.
XX
XX PD 24-OCT-1996.
XX
XX PF 22-APR-1996; 96WO-US05320.
XX
XX PR 07-JUN-1995; 95US-0487429.
XX PR 21-APR-1995; 95US-0426787.
XX PR 07-JUN-1995; 95US-0476102.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (UYJO ) UNIV JOHNS HOPKINS.
XX
XX PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
XX WPI: 1996-485782/48.
XX
XX DR Haemophilus influenzae Rd genome recorded on computer readable
XX medium - useful for identifying commercially important nucleic acid
XX fragments by homology searching
XX
XX PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
XX CC This sequence represents the complete genome sequence of the bacterium
XX CC Haemophilus influenzae strain Rd. The invention relates to a computer
XX CC readable medium (CRM) having recorded upon it the complete H. influenzae
XX CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
XX CC sequence at least 9% identical to (1). By providing the full-length
XX CC genomic sequence in a computer readable form, it is possible to identify
XX CC commercially important nucleic acid fragments and expression modulating
XX CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
XX CC regulate the expression of a nucleic acid molecule. Vectors and altered
XX CC organisms comprising the predicted ORFs can be used to produce any of the
XX CC polypeptide fragments of the H. influenzae Rd genome.
XX
XX SO Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Query Match          14.0%; Score 180; DB 17; Length 1830121;
Best Local Similarity 57.48%; Pred. No. 3.6e-38;
Matches 324; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

OY 348 agccgtcacacgcgtctaccacgcgctgataagtgatgacagagatgtgtgac 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18174 agcgggaaacgattatcccccacagaagatgtatttaacgattcaataatactgct 18233

OY 408 atcaagaatgtgaagaagatgattcttagcgacagacccttaccaggttcccaacga 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18234 ttgagagatgtaaagtgtaattttagtgacagatcccttaacgtgacacaacacagcg 18293

OY 468 catgactcgtttcaagtgtgtgcaaaagcagttccctcccccagctcgtgtgaacata 527

```



```
XX AAF21610;
AC
XX 13-MAR-2001 (first entry)
DT
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
diagnosis; antigen; detection; infection; gene therapy; antibacterial;
ds.
XX
OS Neisseria meningitidis.
XX
XX WO2000066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettein H, Venter JC, Masigiani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
XX WPI: 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
XX used.
XX
SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;
```

```
Query Match 13.4%; Score 172.2; DB 21; Length 349980;
Best Local Similarity 56.7%; Pred. No. 2.3e-36;
Matches 318; Conservative 0; Mismatches 243; Indels 0; Gaps 0;
```

```
QY 352 gtacaccgctctaccgccggtgatcaagtgtacatttgacaggaatgtgtacattc 411
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25875 gacgaatcatctaccgcgcgcgcgtgtgttcaacgaattccgcttcacagcgttcg 25934
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 25935 accggttcaaaagccgtatcttcgtgacaaagatccgtatcaagggtgacggcgacg 25994
QY 472 gactctgttcaagtgtgcaaaagccagttccctcccccagctctgtgaatataca 531
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25995 gtttgatcttccgtccgcgaaggatccgcataccgcgcttctactcaatatactaca 26054
QY 532 aagattgtgtacagacatgatgtgttcaagacatccctggagacatggatataagcgat 591
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26055 aggaattggaaccgaacatcgaagcctttccatcccgcgacaggtctgtgacagcgt 26114
QY 592 gggcaaaacaaagggtgtgtctgtcttaacgcggtgtgacgtgtggtcccatagagcca 651
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26115 gggcgaggaagcaggtgtatgtcttctgaacacaggttttgacggtgtgcaagaaaggcgc 26174
QY 652 actccacaagaagacagaggtctggaagaccttccacagcgtgtgatcaagtgtctgacg 711
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26175 attcgacagccctttagctgtggaacgctttacagatccgtatatacgaagcttgcga 26234
QY 712 tcaacgggaagagaggtgttctctgtgtgtgtggtctatacgccataaaggagagcga 771
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26235 cacacgcgaagcactgtcttcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 26294
QY 772 ccacgacaggaagaacgtacacatgtcttcaagctgttcatcattctcttctgtctgtc 831
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26295 tgatagacagtcgaatcatcatttgatgtgacgcacgcgcatctctctgtcgcgat 26354
QY 832 atcgtgggttctctgtgtgtgtgaagcacttcccaagcttaacggctgaagcgtctgaatctg 891
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26355 atcgcggttcttctgtcgtcgcgcattttcacagcgaacagatatttgccgcgcagc 26414
QY 892 ggacggagcctataactcga 912
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26415 gtatcgacatcgataaactcga 26435
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
AAAX13540/C
ID AAX13540 standard; DNA; 2469 BP.
XX
AC AAX13540;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:603.
XX
KM Enterococcus faecalis; contig; detection; Enterococcal infection;
KM vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
XX WO9805055-A2.
XX
PN 12-NOV-1998.
XX
PD 04-MAY-1998; 98WO-US08985.
XX
PF 14-NOV-1997; 97US-0066009.
XX
PR 06-MAY-1997; 97US-0044031.
XX
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Dillon PJ, Kunsch CA;
XX
XX WPI: 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1; Page 1853-1855; 2084pp; English.
XX
```

CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the *Enterococcus faecalis* genome.
 CC AAX12938 to AAX13919 represent these nucleotide sequences which are
 CC primary nucleotide sequences, also known as contigs. The computer-based
 CC system can identify fragments of the *Enterococcus faecalis* genome with
 CC commercial importance. The products can be used to detect the presence
 CC of *Enterococcus faecalis* in samples. They can also be used for
 CC diagnosing *Enterococcal* infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of *Enterococcus faecalis*, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an *Enterococcal*
 CC infection.

XX Sequence 2469 BP; 771 A; 471 C; 483 G; 738 T; 6 other;

Query Match 12.5%; Score 160.8; DB 20; Length 2469;
 Best Local Similarity 54.2%; Pred. No. 2,7e-34;
 Matches 349; Conservative 0; Mismatches 292; Indels 3; Gaps 1;

QY 268 agacttgagaagagagctgctgcagagcttgaagaagcattactcaacaattgagt 327
 DB 999 ATAGTTGGCAAAACATTATTCTGCTCAATTTGAAAAACCTATTATCAAGATTGGCGG 940
 QY 328 cctttagctgtagtagagagagccgtacacacgtctaccacacgagctgatacga 387
 DB 939 AATTTTGAAGAAAAAGATATCAGACAAACGATTTATCCAGATATATCATATTATTTT 880
 QY 388 gtccgaagagtagtgcagattcaagaatgtgaagaatgagatctctggcgaagacctt 447
 DB 879 CAGGTYAGACTAACCCCTTTTGAAGAGATTAAGATTTTGGGCAAGATCAT 820
 QY 448 accaagctcccaaccaacacacatgactcgttctcagtgagcaaaagccagctccctc 507
 DB 819 ATCATGAGCCCAATCAACGACACAGGACTAAGTTCTCTGTCACACGAGGTGAAGATGC 760
 QY 508 ccccaagctcgtggaacatacaagaatgtgtacacgacatgtagtgcctcaagcattc 567
 DB 759 CGCATCTCTTGGCAATATTTTACAAAGATTTACAAAGCTGA--TTTAGGGTACCAACAG 703
 QY 568 ctgagacatgagatcgaagagagtgagcaaaagaggtgcctgctcctaagcagtgac 627
 DB 702 TCACACATGGCTTTTGAAGAGCTGGGCCAACAAGCGTGTATTATTAAATACCTGTGT 643
 QY 628 tgacgtgagggccatcagcagcaccacccaagaagacagagctgagagacattcacag 687
 DB 642 TGAACGTTGGGCGACGTCAGAGCTATTCCTCATCGAGTTAAAGGTGGGAGCACTGACAG 583
 QY 688 acgctgtgataagtgctgtagcgctcaacggggaaggagctgttccctgtgtgggct 747
 DB 582 ATGTCATTATCGAAGAAATGATGAGCGTGAAGAAACCTGTTGTTTCAATTATTATGAGGAC 523
 QY 748 cataagccataagaagagggcgaccatcgacaggaagacgtcaccatgctcttgaagctg 807
 DB 522 GTCCCTGCTCAAGAAAAATTAAATGATTATAGACTCGGCATGATTAATCAAAATCGC 463
 QY 808 ttcattcatctcttctgctcatcgtgtggttcctgttgaagacattcccaag 867
 DB 462 CACACCCAAAGTCCATTGTACGACATCGTGTTTGGCTTCACGTCGCTTTTCTCAAG 403
 QY 868 ctaagggctgctgaactactctggaagcggagcctataaactgg 911
 DB 402 CGAATGCCGCTCTAGACGTTTAGGGGAACGCCGATTTGATTGG 359

RESULT 12
 AAX27774
 ID AAX27774 standard; DNA; 681 BP.
 XX
 AC AAX27774;
 XX

DT 03-JUN-1999 (first entry)
 XX
 DE B. pallidus uracil DNA glycosylase enzyme (Bpa UDG) encoding DNA.
 XX
 XX Uracil DNA glycosylase: Bpa UDG; enzyme: catalase; hydrolysis:
 KW nucleic acid amplification; thermostable; ss.
 KW
 OS *Bacillus pallidus*.
 XX
 PN US5888795-A.
 XX
 PD 30-MAR-1999.
 XX
 PF 09-SEP-1997; 97US-0926055.
 XX
 PR 09-SEP-1997; 97US-0926055.
 XX
 PA (BECT) BECTON DICKINSON & CO.
 XX
 PI Hamilton PT;
 XX
 DR WPT; 1999-243212/20.
 XX
 DR P-PSDB; AAY01335.
 XX
 PT Nucleic acid encoding uracil DNA glycosylase from *Bacillus pallidus*

Claim 1; Columns 7-12; 7pp; English.

CC This DNA encodes a novel uracil DNA glycosylase enzyme (Bpa UDG) of
 CC *Bacillus pallidus*. Host cells transformed with the vector comprising the
 CC Bpa UDG gene can be used for the recombinant expression of the protein.
 CC The enzyme selectively catalyses hydrolysis of the N-glycosylation link
 CC between uracil and deoxyribose, creating an a basic site that is
 CC susceptible to acid or alkaline hydrolysis. UDG is used to inactivate
 CC contaminating amplicons (containing uracil) in nucleic acid amplification
 CC mixtures, particularly at high temperature. Antibodies raised against UDG
 CC can be used to purify or detect UDG (in usual immunoassays), also to
 CC inactivate it, e.g. as a replacement for, or in combination with, heating
 CC or use of uracil glycosylase inhibitory protein, for control of UDG
 CC activity. The Bpa UDG, which has similar kinetics to the enzyme from
 CC *E. coli* is thermostable, thereby allowing decomposition of contaminating
 CC amplicons at higher temperatures, e.g. 45-75 deg. C than is possible with
 CC known mesophilic enzymes. This reduces the size of temperature changes
 CC between the various stages of the decontamination process.

Sequence 681 BP; 210 A; 128 C; 159 G; 184 T; 0 other;

Query Match 12.4%; Score 159.6; DB 20; Length 681;
 Best Local Similarity 53.8%; Pred. No. 3e-34;
 Matches 352; Conservative 0; Mismatches 299; Indels 3; Gaps 1;

QY 260 ttctggagagacttgagaagagagctgctgcagagtttgaagaagcattactcaaca 319
 DB 12 tttaaaaatgagatggtggggagcgtgttgaagaagagatttgaagaagcattactcaaga 71
 QY 320 atgatactcctttagctgtagtagagagcgctgacacogcttaccacogctgata 379
 DB 72 gctgcgggaatttttgaagcaggaatatagccatcatacagattatccgagatgatacga 131
 QY 380 agtataagctgcagaagagatgtagacattcaagaatgtaagaatgatttcaagcca 439
 DB 132 cattttaagctttagattatacagcctttagaagaatgaagttgctatttaagcca 191
 QY 440 ggaaccttaccaggtcccaaccaagacatgagctgtttagagtgtagcaaaagcagct 499
 DB 192 agatcgtatcatgagccaataatcaacacatgagattagcttttcgtlaaagccgggagat 251
 QY 500 tccccctccccagctcgtgtaacatacaagaatgtagtaccgagatgtgagct 559
 DB 252 tgcacagccgcgctcctgtaaaatatttcatcgagcttgagagaaga---ctcggcgtg 308
 QY 560 caagatcctggagacatgagatcgaagcgaatgtagcaaaaagaagggtgctgctgcttaa 619

Db 309 caagctcccaaacacggtcatcttctcaaatggcgaaagcgggtcttatatgaa 368
 QY 620 cgcgggtgctgacggtgagccatcagccaaactcccaagaagcgtggagac 679
 Db 369 cactgtgctgacggtgaaagcagcagcattctcataaagaagaagcggagaca 428
 QY 680 ctccaccgacgctgtgatacaagtgcgtgacgtcaaccgggaagcgtcttctgt 739
 Db 429 attaccgatacgaagtattagctgtttaacgagcgagagcgtccatcttattt 488
 QY 740 gtgggctcctaagccataaagaaggagcgaccatcgacaggaataacgtacatgctt 799
 Db 489 acggggaagcagcagcagcgaataaagaacatgatacgtacgaatcttgcat 548
 QY 800 gcaagctgtcatcatctctctgtctgctacgtggtgtctctgtgttaagacact 859
 Db 549 tgggtgcgcgcatccatcgccatttctgcaacaagagcgttttttggcagcagcctt 608
 QY 860 ctccaagcctaagcggctgctgaaactatctggagcggagcctataaactggag 913
 Db 609 ctctaaagcaatacattttaagatccatcgatcagagaataatgactggtg 662

RESULT 13

AAQ13285 standard; DNA; 8753 BP.

AAQ13285;

25-OCT-1991 (first entry)

P.denitrificans EcoRI-EcoRI fragment contg. cob f-m genes.

cob gene; corrinoid; descobalocorrinoid; cor gene; ds.

Pseudomonas denitrificans.

Location/Qualifiers

736..1524

/tag- a

/note- "COBF

1620..2999

/tag- b

/product- COBG

3002..3634

/tag- c

/note- "has precorrin-8x mutase activity"

3631..4368

/tag- d

/product- COBI

/note- "has S'-adenosyl-L-methionine:precorrin-2-

4365..5129

/tag- e

/product- COBU

/note- "catalyses transfer of methyl groups"

complement (5108..5892)

/tag- f

/product- COBK

/note- "has precorrin-6x reductase activity"

5862..7103

/tag- g

/product- COBL

7172..7930

/tag- h

/product- COBM

/note- "catalyses transfer of methyl groups"

MO911518-A.

XX

PD 08-AUG-1991.
 XX 30-JAN-1991; 91WO-FR00054.
 PF 31-JAN-1990; 90FR-0001137.
 PR (RHON) RHONE-POULENC BIOCH.
 PA Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schli S;
 PI Thibaut D;
 PI WPI: 1991-252650/34.
 DR P-PSDB; AAR13497-R13501, AAR13505-7.
 DR New polypeptide(s) involved in cobalamin and cobamide
 PT biosynthesis - and DNA encoding them, for amplification of
 PT cobalamin, esp. coenzyme B12 prodn.
 XX
 PS Claim 1; Fig 8; 299pp; French.
 XX
 CC This is a 8.7 kb EcoRI-EcoRI fragment of plasmid pXU367 isolated from
 CC a P.denitrificans genomic DNA bank constructed in vector pXU59.
 CC Plasmids able to complement P.putida and Agrobacterium tumefaciens cob
 CC mutants were identified and their inserts were sequenced. This insert
 CC was found to contain 8 putative open reading frames (one contained
 CC within the complementary strand). The presence of cob genes f-m was
 CC subsequently verified by genetic analysis.
 CC See AAQ13284-Q13288.
 CC
 SQ Sequence 8753 BP; 1450 A; 2772 C; 2954 G; 1577 T; 0 other;

Query Match 11.5%; Score 147.8; DB 12; Length 8753;
 Best Local Similarity 58.7%; Pred. No. 2,2e-30;
 Matches 293; Conservative 0; Mismatches 202; Indels 4; Gaps 2;

QY 415 atgtgaagtagatgattctgagccagacccctaccagcgtcccaacgaacatgac 474
 Db 125 aggtgcgctggtatctgctgcaagcagcctcctcagctgtagcggcggatggc 184
 QY 475 tctgttcaagtgtcaaaagccagctccctccctccctccctccctccctccctccctcc 534
 Db 185 tctgttcaagctgtgcccgcgggtgtccgcagcgcgcgcgtgtgttaacatctacaag 244
 QY 535 aattgtaccgacatgattgtctcaagcattccctgacatgagatcgaagcgtgg 594
 Db 245 aactgaatacga---tctcgatctccgcgcgcgtctcaagcttctcgaagcgtgg 301
 QY 595 caaacaggggt 654
 Db 302 caagcaggggt 360
 QY 655 ccccaagagcagagcgtgtgagaccttcacacgagcgtgtgataaagtgtgagcgtca 714
 Db 361 caacacaggggt 420
 QY 715 acgaggaagagtgcttctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 774
 Db 421 ccgagcagcgt 480
 QY 775 tgcagaggaagcgtacacatgctgtgcaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 834
 Db 481 tgcagcgt 540
 QY 835 gtgggtctctgt 894
 Db 541 ccggttctcgt 600
 QY 895 cgaagcctataaactggag 913
 Db 601 tcatccgactgactggcg 619

RESULT 14
AAC36561
ID AAC36561 standard; DNA; 1348 BP.
XX
AC AAC36561;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 14258.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

```
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151308.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0156559.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.
```

Query Match 11.1%; Score 142.2; DB 21; Length 1348;
Best Local Similarity 54.3%; Pred. No. 3.1e-29;
Matches 310; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

```
QY 341 tggagagagcgctcacacgcgttaccacgcggtgtatcaagtgtacagttcgacagagat 400
DB 483 tgcacgtlaaagccctctgtattatccacgcgcagcactgtatttcaatgtcttataac 542
QY 401 gtgtgacattcaagatgtgaagtagtattctagccagacccttaccagagttcccaa 460
DB 543 aactcctttgtacgaggttaagactgtcaattatcgacaagatcccttatacatgtgacctg 602
QY 461 ccaagacatgagactgtttcagttgtgcaaaagccagttcccccctcccccagttcgt 520
DB 603 tcaagctatggtgtttgtctctctgtactgtgagagagaaagcttccctctagtctgtt 662
```

```
QY 521 gaacatacaagaattgtgtacccagacattgttggtcttcaagcattccgtgacatgagaga 580
DB 663 gaacattttaagagattcataaagaatgtgtgtcttccatcccaac---gtcacgtaa 719
QY 581 tctaagcggatggtgcaaaaaggggtgtcgtgtcttaagcgggtgtcgtacggtgtgggc 640
DB 720 tctacagaatggtgctgtgcaagggtgtgttactccctgaatgtgtcttctacaglaaggag 779
QY 641 ccataggcccaactcccaagaagacagaggtgtggtgagaccttaccagagcgtgtgataa 700
DB 780 taacagcctaattcataatcatgcaagaagagatgtggaaacaattcacgtatgtatttca 839
QY 701 gtgtcgtgagcgttcaaccgggaaggagtgcttctcctgtgtgtgtggtgtacagccataa 760
DB 840 aaglatcttcacagcaagaaggaagggtgtgttcttctctctcgtgggaagatacgtcaaga 839
QY 761 gaaggagcgaccattcgacagagaagcaagcgtacatgtcttgcaagcgtttatcatcttc 820
DB 900 gaatccaagtgtatagatgtcgactaaacatacatatactacagcagctatccacttgg 959
QY 821 ttgtctgtcatcgttgtgtcttctgtgtgttaagcacttccaaaggtacagggctgtct 880
DB 960 ttgtcgcgcaatagagagcttcttcgacttcgagcattctctcgcgcgaaccagctact 1019
QY 881 gaaactatctggagcggagcctataactgg 911
DB 1020 cgaggaaatggtggtatctcccatagactgg 1050
```

RESULT 15

ID ABA03041 standard; DNA: 2944528 BP.

AC ABA03041;

DT 05-FEB-2002 (first entry)

DE *Listeria monocytogenes* EGD-e genome sequence.

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

KW vitamin B12; bacterial infection; disease; ds.

OS *Listeria monocytogenes*.

PN W020017735-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001MO-FR01118.

PX 11-APR-2000; 2000PR-0004629.

PA (INSP) INSP PASTEUR.

XX Buchrieser C, Frangeul L, Couve E, Rusnock C, Fsihi H, Dehoux P;
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kretz J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;

XX MPI: 2002-010914/01.

XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
PT and prevention of *Listeria* and related bacterial infections, and
PT related polypeptides

XX Claim 1; SEQ ID No 1; 192pp; French.

CC The present sequence is the genome sequence of *Listeria monocytogenes*
CC EGD-e. This sequence and fragments of this sequence are useful for

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2002, 03:37:10 : Search time 83.01 Seconds

(without alignments)
3796.504 Million cell updates/sec

Title: US-09-758-017A-1

1283

Sequence: 1 gacatccgcgtgcacaatg.....agaataataatatttta 1283

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/lna/5A_COMB.seq: *
2: /cgn2_6/prodata/2/lna/5B_COMB.seq: *
3: /cgn2_6/prodata/2/lna/6A_COMB.seq: *
4: /cgn2_6/prodata/2/lna/6B_COMB.seq: *
5: /cgn2_6/prodata/2/lna/PCUTUS_COMB.seq: *
6: /cgn2_6/prodata/2/lna/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	207.6	16.2	1532	4	US-09-651-656-36
2	107	8.3	35100	2	US-08-770-379-18
3	107	8.3	35100	4	US-08-757-669A-18
4	107	8.3	35100	4	US-09-230-371A-18
5	89	6.9	4403765	4	US-09-103-840A-2
6	89	6.9	4411529	4	US-09-103-840A-1
7	48.2	3.8	7218	1	US-08-232-463-14
8	46.4	3.6	7218	1	US-08-232-463-14
9	41.2	3.2	19124	2	US-08-487-826B-13
10	39.4	3.1	72604	4	US-09-268-992-7
11	37.4	2.9	1689	1	US-07-991-867B-41
12	37.4	2.9	1689	2	US-08-544-332-41
13	37.4	2.9	8457	1	US-07-991-867B-1
14	37.4	2.9	8457	2	US-08-544-332-1
15	37	2.9	1850	3	US-08-617-860B-32
16	37	2.9	4098	2	US-08-605-106-4
17	36.2	2.8	636	4	US-08-998-416-1137
18	36.2	2.8	837	4	US-08-998-416-288
19	36.2	2.8	1411	3	US-09-188-930-12
20	36	2.8	2223	1	US-08-257-073-4
21	35.8	2.8	319	1	US-07-593-657-14
22	35.4	2.8	678	1	US-07-991-867B-23
23	35.4	2.8	678	1	US-08-107-755A-23
24	35.4	2.8	678	2	US-08-544-332-23
25	35.4	2.8	4818	3	US-08-817-926-27
26	35.4	2.8	6768	1	US-08-107-755A-1
27	35.4	2.8	8457	1	US-07-991-867B-1

C	28	35.4	2.8	8457	2	US-08-544-332-1	Sequence 1, Appl
C	29	34.6	2.7	615	4	US-08-998-416-186	Sequence 186, App
C	30	34.6	2.7	6519	1	US-08-233-008A-7	Sequence 7, Appl
C	31	34.4	2.7	826	2	US-08-238-821B-51	Sequence 51, Appl
C	32	34.4	2.7	826	2	US-08-238-821B-51	Sequence 61, Appl
C	33	34.4	2.7	826	5	PCT-US95-05744-51	Sequence 51, Appl
C	34	34.4	2.7	826	5	PCT-US95-05744-51	Sequence 61, Appl
C	35	34.2	2.7	5181	1	US-08-257-073-10	Sequence 10, Appl
C	36	34.2	2.7	5703	4	US-09-056-075-1	Sequence 36, Appl
C	37	34.2	2.7	6243	2	US-09-056-075-1	Sequence 36, Appl
C	38	34	2.7	3108	4	US-08-968-752B-3	Sequence 3, Appl
C	39	34	2.7	3114	4	US-09-107-149-18	Sequence 18, Appl
C	40	34	2.7	3222	3	US-08-968-752B-1	Sequence 1, Appl
C	41	34	2.7	3852	1	US-07-867-106-2	Sequence 2, Appl
C	42	34	2.7	9048	3	US-08-973-273-4	Sequence 4, Appl
C	43	34	2.7	11225	6	5182210-9	Patent No. 5182210
C	44	34	2.7	51952	3	US-08-947-823-1	Sequence 1, Appl
C	45	34	2.7	246240	2	US-08-724-394A-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1									
US-09-651-656-36									
; Sequence 36, Application US/09651656									
; Patent No. 6340566									
; GENERAL INFORMATION:									
; APPLICANT: MCCUTHEN-MALONEY, SANDRA									
; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE									
; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,									
; FILE REFERENCE: IL-10689									
; CURRENT APPLICATION NUMBER: US/09/651,656									
; CURRENT FILING DATE: 2000-08-29									
; PRIOR APPLICATION NUMBER: 60/192,764									
; PRIOR FILING DATE: 2000-03-28									
; NUMBER OF SEQ ID NOS: 106									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 36									
; LENGTH: 1532									
; TYPE: DNA									
; ORGANISM: Escherichia coli									
US-09-651-656-36									
Query Match 16.2%; Score 207.6; DB 4; Length 1532;									
Best Local Similarity 57.9%; Pred. NO. 1.1e-53;									
Matches 388; Conservative 0; Mismatches 279; Indels 3; Gaps 1;									
QY	258	ggttcggaagacttgagaagagactgctgcagacttgcgaagacttc	314						
DB	536	gctaacgaataacactgcatgacgtgctgcgaagagaacgaacactatttc	595						
QY	315	aaacaattgcttccttgatgctgataagagagccgtccacacgctcaccacggt	374						
DB	596	ataacacctcagacgctgcagcagcgagcgagcgcgtcactatctaccacacaa	655						
QY	375	gatacagtgtaacggttgacaagagatggtgacattcaaatgtgaagagatgata	434						
DB	656	aaagatgctcttaacggttcgcgtcttaacagagttgggtgaagtgatgattctc	715						
QY	435	ggcagagaccttaccacggttcccaacagacacatgactctgttcagtgcaaaag	494						
DB	716	ggcagagatctctatacagcagcagcgagcagcgatgctgttcgatttcgtgcc	775						
QY	495	ccagttccctcccccagctcgtgatacatatacaagaattgtgtacagcatgat	554						
DB	776	ggatctcagatctctcgtcgtatattatgataatgagactggaaatactattccg	835						
QY	555	ggcttcagatctctgacatgagatcgaagcgatggcgaacaaaggggtgctgctg	614						

Db 836 ggcctaccgcccataatcgtgtatcttgaaagctggcgctgcaagcgcttctgcta 895
 QY 615 cttacacggtgtcgtacgctgacgagccatccacccacgaagagagcgctgg 674
 Db 896 ctcaatctgtgttgagagtgacgagctgacgagctgacgagctgacgagctgag 955
 QY 675 ggaaccttcacgagcgtgtgacatgaagctgagcgtgacacgaggaagagcttctc 734
 Db 956 gaaaccttcacgagtaagtgatcagcgtgattacacgacatcgagagcggtgtt 1015
 QY 735 ctgttgaggagcctacacgacataaagaagggagcaccatcgacagaagacgaccat 794
 Db 1016 ttgttgaggagcctacacgacataaagaagggagcaccatcgacagaagacgaccat 1075
 QY 795 gcttcgaagcgtgtcgtacatcctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 854
 Db 1076 gtaactgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1135
 QY 855 caacttcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 914
 Db 1136 catttgctgagcaaatcagtgctgagcaacgctgagcagcagcagcagcagcagc 1195
 QY 915 gcactctaac 924
 Db 1196 ccagctatc 1205

RESULT 2

US-08-770-379-18/c
 ; Sequence 18, Application US/08770379
 ; Patent No. 5849564

GENERAL INFORMATION:

; APPLICANT: Chang, Yuan
 ; APPLICANT: Bohenzky, Roy A.
 ; APPLICANT: Russo, James J.
 ; APPLICANT: Edelman, Isidore S.
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
 ; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/770,379
 ; FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 52342
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 35100 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-770-379-18

Query Match 8.3%; Score 107; DB 2; Length 35100;
 Best Local Similarity 52.8%; Pred. No. 6.1e-22;
 Matches 255; Conservative 0; Mismatches 225; Indels 3; Gaps 1;

QY 356 caccgttaccaccgagctatcaagtgtacagttcagacagagatgtgtacattcaaga 415
 Db 34128 CACAGTATATCCCCACAGAGCAAGCTGATGTGTGTCCATGCTGCATCCAGAGGA 34069
 QY 416 tgtgaagatgtatgtatcctaagccagagacccttaccagcagctcccaacgaacagact 475
 Db 34068 TATTAAAGTGTATCTTAAAGCCAGAGACCCGTACACAAAGG---CCAAGTACTGGCCT 34012
 QY 476 ctgttcaagtgtgcaaaagccagtlccctcccccagctcgtgtgaacalatacaaga 535
 Db 34011 GGCGTTAAGTGTGATCCGCAATGTCCAGCTTCCAGCTTGAAGCATCTTTAGAGA 33952
 QY 536 atgtgtaccgacattatgtgtcctcaagcaltccctgagacatgagagatcaagcgatggc 595
 Db 33951 GCTAGAGGCTTCCGCTCCCAATTTCACTCTTCCACGGGTGCTCCGACAGCTGGGC 33892
 QY 596 aaacaaggggt 655
 Db 33891 TCGCAGAGGT 33832
 QY 656 ccacaagagacagagctgtgagaccttcaccgagcgtgtgtatcaagtgtgtgtgtgtgtgtgt 715
 Db 33831 ACACAGAGGAGCTTGGCTGGAGT 33772
 QY 716 ccggaagagagctgt 775
 Db 33771 GTTAGAATGATGCTTTTCTCTGT 33712
 QY 776 cgacaagaaagcgtacatgtcttgcagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 835
 Db 33711 AAACGACAGAAACACCTGATGT 33652
 QY 836 tgg 838
 Db 33651 TGG 33649

RESULT 3

US-08-757-669A-18/c
 ; Sequence 18, Application US/08757669A
 ; Patent No. 6183751

GENERAL INFORMATION:

; APPLICANT: Chang, Yuan
 ; APPLICANT: Bohenzky, Roy A.
 ; APPLICANT: Russo, James J.
 ; APPLICANT: Edelman, Isidore S.
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
 ; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/757,669A
 ; FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-18

Query Match 8.3%; Score 107; DB 4; Length 35100;
Best Local Similarity 52.8%; Pred. No. 6.1e-22;
Matches 255; Conservative 0; Mismatches 225; Indels 3; Gaps 1;
QY 356 caccgctaccaccgagctatcaagtgtacagttcgacagagatgtgtacattcaaga 415
DB 34128 CACAGGTATCCCCACAGACAGCTGATGTGTGCTCCACTGCTGCATCCAGAGGA 34069
QY 416 tggaaagtgtatcttaagccagagacccttaaccaggtcccaaccacacatgagct 475
DB 34068 TATTAAAGTGTATCTTAGGCCAGACCCGTACACAGAAGG---CCAACTACTGGCTT 34012
QY 476 ctgttcagtgtgcaaaagccagttcccccctcccccagctcgtgacacatacaaga 535
DB 34011 GCGGTTAGTGTGATCCGCAATGTCAAGTTCACCCAGTTTGAGAAAGCATCTTTAGAGA 33952
QY 536 atgtgtaccgaatgtatgagcttcaagcactctgagacatgagatctaaagcagatggc 595
DB 33951 GCTAGAGGCTTCGTCCTCCCAATTTCAGTACTCTTCCACAGGGGTGCTCGACAGCTGGGC 33892
QY 596 aaaaagaggtgtcgtctgtcttaacgcggtgtcagcgtgacggccatcagcgaactc 655
DB 33891 TCGCCAGGGTGTGTGCTTACTTAAACACAGTTTTCACAGATTTCATCATCATATCTTCAA 33772
QY 656 ccacaaggaagaggtgtggaagaccttcaaccgagcgtgtgataagtgtgagcgtcaa 715
DB 33831 ACACGAGGAGCTTGGCTGGGATTTGTTTCACAGATTTCATCATCATATCTTCAA 33772
QY 716 ccgggaagaggtgttctcctgtgtgtgagctacatagccacataagaagagcgacat 775
DB 33771 GTTAGAACATTTGGCTTTTCTCTCGTGGGGCCGCAAGGCGATTGACAGAACTCGCTCAT 33712
QY 776 cgacaggaagagcgtacacatgtcttgaagctgttcatccatctcctgtgtcgtacgc 835
DB 33711 AAACGACACAAACCTGTGTGCTTACGGCCACAGATCATCTCGCTGGCTCTCTTGG 33652
QY 836 tgg 838
DB 33651 TGG 33649

RESULT 4
US-09-230-371A-18/c
Sequence 18, Application US/09230371A
Patent No. 6348586

GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A
APPLICANT: Russo, James J
APPLICANT: Edelman, Isidore S
APPLICANT: Moore, Patrick S
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: 45185-G-PCT-US
CURRENT APPLICATION NUMBER: US/09/230,371A
CURRENT FILING DATE: 1999-11-17
PRIOR APPLICATION NUMBER: PCT/US97/13346
PRIOR FILING DATE: 1997-07-22
NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 18
LENGTH: 35100
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-18

Query Match 8.3%; Score 107; DB 4; Length 35100;
Best Local Similarity 52.8%; Pred. No. 6.1e-22;
Matches 255; Conservative 0; Mismatches 225; Indels 3; Gaps 1;

QY 356 caccgctaccaccgagctatcaagtgtacagttcgacagagatgtgtacattcaaga 415
DB 34128 CACAGGTATCCCCACAGACAGCTGATGTGTGCTCCACTGCTGCATCCAGAGGA 34069
QY 416 tggaaagtgtatcttaagccagagacccttaaccaggtcccaaccacacatgagct 475
DB 34068 TATTAAAGTGTATCTTAGGCCAGACCCGTACACAGAAGG---CCAACTACTGGCTT 34012
QY 476 ctgttcagtgtgcaaaagccagttcccccctcccccagctcgtgacacatacaaga 535
DB 34011 GCGGTTAGTGTGATCCGCAATGTCAAGTTCACCCAGTTTGAGAAAGCATCTTTAGAGA 33952
QY 536 atgtgtaccgaatgtatgagcttcaagcactctgagacatgagatctaaagcagatggc 595
DB 33951 GCTAGAGGCTTCGTCCTCCCAATTTCAGTACTCTTCCACAGGGGTGCTCGACAGCTGGGC 33892
QY 596 aaaaagaggtgtcgtctgtcttaacgcggtgtcagcgtgacggccatcagcgaactc 655
DB 33891 TCGCCAGGGTGTGTGCTTACTTAAACACAGTTTTCACAGATTTCATCATCATATCTTCAA 33772
QY 716 ccgggaagaggtgttctcctgtgtgtgagctacatagccacataagaagagcgacat 775
DB 33771 GTTAGAACATTTGGCTTTTCTCTCGTGGGGCCGCAAGGCGATTGACAGAACTCGCTCAT 33712
QY 776 cgacaggaagagcgtacacatgtcttgaagctgttcatccatctcctgtgtcgtacgc 835
DB 33711 AAACGACACAAACCTGTGTGCTTACGGCCACAGATCATCTCGCTGGCTCTCTTGG 33652
QY 836 tgg 838
DB 33651 TGG 33649

RESULT 5
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328

GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g


```

; TELEFAX: (703)683-4109
;
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZ9pt-Fls
US-08-232-463-14

Query Match      3.8%; Score 48.2; DB 1; Length 7218;
Best Local Similarity 9.9%; Pred. No. 0.0025;
Matches 38; Conservative 181; Mismatches 164; Indels 0; Gaps 0;

QY 788 tcacacatgtcttgcagacttctcattcattccttctgtcgtccatcggtggtccttg 847
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1128 yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1187

QY 848 ttgtaagcactctccaagcctcagcgctcgtgaactatctggagcgagccataaa 907
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1188 yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1247

QY 908 ctggagagcacttaactcttctgctccttaactgcttaactgttttaagatgaacat 967
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1248 yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1307

QY 968 cacactatcttctacagcttctccaagttcaacacatctataacttcttcttct 1027
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1308 yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1367

QY 1028 ttggaatgtagctgtcttctgctcgtttagataactaaacacttaccactcgca 1087
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1368 yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1427

QY 1088 tcttgactactgttcagcataacttccacaacttgacaaaatgtatttataa 1147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1428 yyyyyyygTACCAATTCCTATCTCTTACTTACTGATGATGATATATACAGTG 1487

QY 1148 ttgatatctctgtatcataaa 1170
   || || || || || || || || || || || || || || || || || || || || ||
Db 1488 ATGCTACTATGCCGTTTGTGAA 1510

RESULT      8
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
;
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZ9pt-Fls
US-08-232-463-14

Query Match      3.6%; Score 46.4; DB 1; Length 7218;
Best Local Similarity 4.7%; Pred. No. 0.00086;
Matches 17; Conservative 196; Mismatches 147; Indels 0; Gaps 0;

QY 108 aaatgaatgaagaacgccaagaactgagtcctcaaatgtggaacaaagactca 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1391 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1332

QY 168 tcgcaagcttctagtggaagcgttgaaagatcgccaaataaagaagacgcctt 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1331 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1272

QY 228 gacaagattgagaaagcagccctgcaggttcggaagacttggaagagaagctg 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1271 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1212

QY 288 gctcagagtttgaagccacttcaacaattgattcttctgttagctgataagag 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1211 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1152

QY 348 agccgtcacaccgctctccacgcgctgatacagtgtaagctgcagagatgtgac 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1151 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1092

QY 408 attcaagatgtgaagtagtattctagccagagacccctacacggtcccaacaagca 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1091 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1032

RESULT      9
US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chltnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas F.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
```



```
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI14.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1689 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-991-867B-41
```

```
Query Match 2.9%; Score 37.4; DB 1; Length 1689;
Best Local Similarity 46.3%; Pred. No. 0.2;
Matches 157; Conservative 0; Mismatches 181; Indels 1; Gaps 1;
```

```
QY 946 ttaactgtttaagatgaacatcacatatttctacaagctttcccaagttcaacca 1005
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 113 TTTCCTTTTACCATATATTTTTCATATATATTCGTTTATTCATTCATCTCTA 172
QY 1006 atctataagcttcattgctcttttggaatgagtcgctcttttgctggttttagagact 1065
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 173 GTTTTAAAGTTTGGCTAGAAATTTCAATATTAAGAAATTAACAAATTAATAATTAACAT 232
QY 1066 aaacaccttacacactgcacatgctgact-cagtctcagtcacaataaactcacacact 1124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233 TTTCCTTTTACCATATATTTTTCATATATTTTATATTCCTTTCCTCTC 292
QY 1125 tgaacaaatgttatttataatgatactatctgacattaaagattgtttttcc 1184
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 293 GAATTTAAATTAATTTTACAAAGTTTCTATATATTTTACAGCAATTTGATTAATATATTA 352
QY 1185 caggcgtcttcacagtagtactagatataaactgatttaacttaacttaactcagatgca 1244
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 353 TAATTAATTTATACAGTGAATGGTAATTAATGTTTATTTTATTAATTAATTAATTAATAA 412
QY 1245 actgcttaagttttatcgagaataaataattatattta 1283
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 413 CTTTCAATTAAGATTCAGAGAAATTAATATATTTTGTGA 451
```

```
RESULT 12
US-08-544-332-41
; Sequence 41, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/544,332
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: UFI14.C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1689 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-544-332-41
```

```
Query Match 2.9%; Score 37.4; DB 2; Length 1689;
Best Local Similarity 46.3%; Pred. No. 0.2;
Matches 157; Conservative 0; Mismatches 181; Indels 1; Gaps 1;
```

```
QY 946 ttaactgtttaagatgaacatcacatatttctacaagctttcccaagttcaacca 1005
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 113 TTTCCTTTTACCATATATTTTTCATATATTTTATATTCGTTTATTCATTCATCTCTA 172
QY 1006 atctataagcttcattgctcttttggaatgagtcgctcttttgctggttttagagact 1065
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 173 GTTTTAAAGTTTGGCTAGAAATTTCAATATTAAGAAATTAACAAATTAATAATTAACAT 232
QY 1066 aaacaccttacacactgcacatgctgact-cagtctcagtcacaataaactcacacact 1124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233 TTTCCTTTTACCATATATTTTTCATATATTTTATATTCCTTTCCTCTC 292
QY 1125 tgaacaaatgttatttataatgatactatctgacattaaagattgtttttcc 1184
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 293 GAATTTAAATTAATTTTACAAAGTTTCTATATTTTACAGCAATTTGATTAATATATTA 352
QY 1185 caggcgtcttcacagtagtactagatataaactgatttaacttaacttaactcagatgca 1244
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 353 TAATTAATTTATACAGTGAATGGTAATTAATGTTTATTTTATTAATTAATTAATTAATAA 412
QY 1245 actgcttaagttttatcgagaataaataattatattta 1283
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 413 CTTTCAATTAAGATTCAGAGAAATTAATATATTTTGTGA 451
```

```
RESULT 13
US-07-991-867B-1
; Sequence 1, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
```

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,860B
FILING DATE: 01-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP94/02950
FILING DATE: 05-SEP-1994
APPLICATION NUMBER: DE P4329951.2
FILING DATE: 04-SEP-1993
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1850 Base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double stranded
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Cuphea lanceolata

```

```

;
; CLONE: CITEg1
;
; FEATURE:
; NAME/KEY: CAAT-Signal

```

```

?      FEATURE:
?      NAME/KEY:  TATA-Signal
?      LOCATION:  1553..1556
?      FEATURE:
?      NAME/KEY:  Transcription start
?      LOCATION:  1585
?      FEATURE:
?      NAME/KEY:  Legumibox
?      LOCATION:  1642..1657
?      FEATURE:
?      NAME/KEY:  Startcodon
?      LOCATION:  1797..1799
?      FEATURE:
?      NAME/KEY:  CDS
?      LOCATION:  1797..1850
?      US-08-617-860B-32

```

	Query Match	2.9%	Score 37	DB 3	Length 1850
	Best Local Similarity	47.6%	Pred. No. 0.26		
	Matches 109	Conservative 0	Mismatches 120	Indels 0	Gaps 0
QY	1054	ttttagaacttaaaacacttaccacctcgcacatgctgacatcagtcctcagcaatataa	1113		
Db	194	TTTTGAATTTTAAATAATTTTTTAAATTTAAATATTTTAAATTCCTTTAAAAAAA	253		
QY	1114	cttcacacactgacaaaatagttatcttctaattgattatcctcgtacataagat	1173		
Db	254	TATTTTAAATATTTATTAATAATTTAGTTTTTAAATTTTTTAAATATTTTAAATTTAGTTT	313		
QY	1174	tgtttttcccaagcgtttccatcagagctacagataacgttataacctaattt	1233		
Db	314	TTTTATTTTAAAAATATGTTGAATTTTTTAAATAATTTTTTGCTTTAAAAAATATATTTA	373		
QY	1234	ccatgagatgcactgccttaagttttttagcgcgaataataatcataattt	1282		
Db	374	AAAGTTTTTAAATATTTTTTGAATTTTTTGAATAATTTTGAATAATTTTGT	422		

Search completed: August 22, 2002, 06:06:53
Job time: 8983 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2002, 02:09:50 ; Search time 1887.91 Seconds

(without alignments)
9172.362 Million cell updates/sec

Title: us-09-758-017a-1

Perfect score: 1283

Sequence: 1 gacatccgcttgcacaaatg.....agaataaataataattta 1283

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST:*

- 1: em_estbta:*
- 2: em_estbta:*
- 3: em_estbta:*
- 4: em_estbta:*
- 5: em_estbta:*
- 6: em_estbta:*
- 7: em_estbta:*
- 8: em_estbta:*
- 9: em_estbta:*
- 10: em_estbta:*
- 11: em_estbta:*
- 12: em_estbta:*
- 13: em_estbta:*
- 14: em_estbta:*
- 15: em_estbta:*
- 16: em_estbta:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	381	29.7	696	9	AJ395572
2	340.2	26.5	1175	10	BM471137
3	317.8	24.8	822	10	BI823926
4	313.8	24.5	808	10	BE236591
5	309	24.1	869	10	BE105781
6	307.6	24.0	609	9	AU126319
7	298.4	23.3	881	10	BM449708
8	295.4	23.0	622	10	BI066406
9	293.4	22.9	792	9	AJ394089
10	284.6	22.2	548	9	AA290918
11	280.2	21.8	863	9	AL559968
12	274.4	21.4	788	10	BI651753
13	272.6	21.2	574	10	BI940552
14	267.2	20.8	768	10	BI657579
15	262.4	20.5	614	10	BE866388
16	260.8	20.3	1061	10	BM458045
17	259	20.2	731	9	AV705903

18	253	19.7	612	10	BI064298
19	250	19.5	672	10	BM104057
20	243.8	18.0	924	10	BE176725
21	243	18.9	874	10	BE099178
22	241	18.8	627	10	BE309210
23	233.2	18.2	535	9	AA068344
24	230.4	18.0	806	9	AL669156
25	229.6	17.9	555	10	H09366
26	223	17.4	407	9	AA401453
27	222.4	17.3	688	10	BI937645
28	217.6	17.0	484	10	BI064543
29	216.4	16.9	491	9	AA633967
30	214	16.7	531	9	AJ396948
31	213.4	16.6	915	10	BE793197
32	212.6	16.6	914	10	BE902908
33	209.8	16.4	539	10	BI979825
34	208.6	16.3	887	10	BM012071
35	207.4	16.2	654	10	BM103222
36	207.4	16.2	938	10	BE389571
37	205.6	16.0	498	9	AA068321
38	204.6	15.9	858	10	BE701762
39	202.4	15.8	629	10	BI691535
40	199.6	15.6	555	10	T78215
41	198.6	15.5	475	10	BE751706
42	197	15.4	621	10	BI980431
43	186.8	14.6	374	9	AA356048
44	186.6	14.5	403	9	AV706222
45	183.4	14.3	532	10	BI675591

ALIGNMENTS

RESULT 1

LOCUS AJ395572 696 bp mRNA linear EST 25-JAN-2001

DEFINITION AJ395572 dktfz426 Gallus gallus cDNA clone 22p14f1, mRNA sequence.

ACCESSION AJ395572

VERSION AJ395572.1 GI:7126625

KEYWORDS EST.

SOURCE chicken.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 696)

AUTHORS Abdulkhmanov, I., Lodygin, D., Gerolt, P., Arakawa, H., Law, A., Plachy, J., Korn, B. and Buerstedde, J.M.

TITLE A large database of chicken bursal ESTs as a resource for the analysis of vertebrate gene function

JOURNAL Genome Res. 10 (12), 2062-2069 (2000)

MEDLINE 20568495

COMMENT Contact: Buerstedde JM

Cellular Immunology

Heinrich-Pette-Institute

Martinistr. 52, 20251 Hamburg, Germany

Email: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.

FEATURES

source location/Qualifiers

1..696

/organism="Gallus gallus"

/strain="CB"

/db_xref="taxon:9031"

/clone="22p14f1"

/clone_lib="dktfz426"

/tissue-type="Bursa of Fabricius"

BASE COUNT 156 a 187 c 205 g 146 t 2 others

ORIGIN

Query Match 29.7%; Score 381; DB 9; Length 696;

Best Local Similarity 72.9%; Pred. No. 3.7e-83;

Matches 489; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

Dd	332	CgGTTcGTGGCCATCAAGCCAACCTCTCATTAAGAGCAGAGCGTGGAGCAGATTTCACGTATG	391
Oy	691	cctgatcaagtgcgtcgagcgtccaacccggaaaggagtcgatttcccttgctgtggagcatat	750
Dd	392	CAGTTGTGTGCCCTGGCTAAATCAGAACTCGAATGCGCTTGTTTCTTTGCTCTGCGGCCTCTT	451
Oy	751	accgccataagaaggagcgacccatcgacaaggaacacgtccaccatgctcttgcaagctgttc	810
Dd	452	ATGCTCAGAGAAGAGGCGACGTGCATTGATAGAGAACCGGACCACCATGTACTACAGACGGCTC	511
Oy	811	atccatctccctttgtctgtcatatcgtygggttcccttgctgtgtaagaccttcccaggcta	870
Dd	512	ATCCCTCCCTCCCTTGTGCAGTGTAAAGGGTTTTTGTGGATGTGAACAATTTTCAAAGACCA	571
Oy	871	accggcgtcgtagaacctactctggagcagagccatatcaactacgtggagcactctcaa	923
Dd	572	ATGACTGTCTGCAGAGTCTGGCAGAAAGCCCATTTGACTGTGAAGAGCTGTGA	624
RESULT	4		
LOCUS	BF236591	808 bp	mRNA linear EST 14-NOV-2000
DEFINITION	60202877121 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4163920 5'		
ACCESSION	BF236591		
VERSION	BF236591.1		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	NIH-MGC http://mgc.ncl.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-remail.nih.gov		
	Tissue Procurement: Jeffrey E. Green, M.D.		
	CNSA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LMNL at:		
	http://image.lnl.gov		
	Plate: LLM9448 row: n column: 17		
	High quality sequence stop: 708.		
FEATURES	Location/Qualifiers		
Source	1..808		
	/organism="Mus musculus"		
	/strain="FVB/N"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:4163920"		
	/clone_lib="NCI_CGAP_L19"		
	/lab_host="DH10B (TI phage-resistant)"		
	/note="Organ: liver; Vector: pCMV-Sport6; site_1: NotI;		
	site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.		
	Average insert size 1.9 kb. Constructed by Life		
	Technologies. Note: this is a NCI_CGAP Library."		
BASE COUNT	191 a 217 c 207 g 193 t		
ORIGIN			
Query Match	24.5%; Score 313.8; DB 10; Length 808;		
Best Local Similarity	69.5%; Pred. No. 1.4e-66;		
Matches	426; Conservative 0; Mismatches 187; Indels 0; Gaps 0;		
Oy	312	tccaacaattgattgctcttgtagtcgtagtagagagcgcgtcacaccgctctaccaccg	371
Dd	111	TTAAACCCAGCTAATGGATTGTGCTGTAAGAAAGAACACACCAAGCTCATCCGCC	170
Oy	372	gctgtcaagtgtacagtgcgacagaagtggtgacatcacaagtgtgaagtgtgatt	431
Dd	171	CCGGGACAGGgtttTACATGAGCACCGAGAtgtCGCACTCCGAGATGTGAAGGTGTCTATT	230

QY	432	ctgaagccaagacccttaccacggtccccaacccaagcacatggaactgtgttcagtgtaa	491
Db	231	ctgggacacagagatccttataccgacctaatcaagctcacgggctctgctttacatgtccaa	290
QY	492	aagcagatccccctcccccaagctcgtgtaacatacaaaaagtgtgtacacgaatt	551
Db	291	agaccagttccaccccccccgccagtttggaaaaaatTTTTTAAAGAGCTGCTCACAGACATC	350
QY	552	gattgcttcaaacatccttggacatgtgagatctcaagcgaatggtgcaaaacaaggtgtcgtg	611
Db	351	gattgttttgtttcatctcgtggccatggggatyytgtcaggggtggcccgacacagtggtcttc	410
QY	612	ctgcttaacgcggtgtcgcacgttcggyggccatacgaagccaactccccaaggaacgaagc	671
Db	411	cttcctcaacggccgctccttgccttgcggcgccacccaagccaatttcccatTAAGGAGAGGGC	470
QY	672	tgggagaccttcaacggaagctgtgatcaagtgtgcgcgtcaaccgggaagagatcgtt	731
Db	471	tgggagcagattcAGGAGCCGGGTGCTGTCCTGGCTGAATCAGAACCTGAGTGGCCTCGTC	530
QY	732	ttccgctgttgggggtctacaacgcccataaagaaggagcgacatcgacaggaacgtcac	791
Db	531	ttccctccttgggggcttattatgcttcagaaagagggcagctgattgacaggaacgctcac	590
QY	792	catgtctgcgaagctgttcatccatccctcttctgtctgcatacgtgtggtcttccttggtt	851
Db	591	catgtctgcagacagcttccacccctccgctgctgctgtacagaggggtcttcgggattgc	650
QY	852	aagcacttcccaaggaagcgaaggtgtcgtcgaacctatctggagcgaagccttcaactgtg	911
Db	651	agacatttttttcaaaagccaatgaatgcttccgaagaagcttgccgaagaagcccatcaactgg	710
QY	912	agagcacttcaac	924
Db	711	AAGGAGGTGTGAC	723
RESULT	5		
LOCUS	Bg105781	869 bp	mRNA linear EST 30-JAN-2001
DEFINITION	602312622p1 NIH_MGC_84 Homo sapiens	CDNA clone IMAGE:4422292 5',	
ACCESSION	Bg105781		
VERSION	Bg105781.1	GI:12599627	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 869)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-rt@mail.nih.gov Tissue Procurement: AROCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: LLM10162 row: 1 column: 05 High quality sequence stop: 655. Location/Qualifiers 1. 869 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4422292" /clone_1id="NIH_MGC_84" /tissue_type="adrenal cortex carcinoma, cell line" /lab_host="DH10B (phage-resistant)"		
FEATURES	source		

BASE COUNT	ORIGIN
221 a	/note="Organ: adrenal gland; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1,229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."
200 c	225 g 223 t

Query Match	24.1%	Score 309	DB 10	Length 869
Best Local Similarity	71.2%	Pred. No. 2,1e+65		
Matches 422	Conservative 0	Mismatches 170	Indels 1	Gaps 1
QY 331	ttgtgtcgtatgagagagagccgctcacaccgctctaccaccacgycgtatcaagtgtacagt	390		
Db 21	TTGTGTCAGAGAAAGAAAGACATTACACTGTTTATCCACCCCCACAGAGTCTCACCT	80		
QY 391	cgaacagatgtgtgacatctcaagaatgtgaaagtgtatcttaggccagaacctacc	450		
Db 81	GGACCCAAATGTTGTACATAAAAGATGGAAGGTTGTATCCTGGGACAGGATCCATATC	140		
QY 451	acgggcccaacaagaacacatgacatcgtcttccagtgtcgaagaacagatccccctccc	510		
Db 141	ATGGACCTCAATCAACACTACCGGCTCTCTTATAGTGTCAAAAGCCCTTTCGCTCCGC	200		
QY 511	ccagtcctcgtgaaacatalaacaagaattgtgtaccgcacattgtatggtctcaagaacctg	570		
Db 201	CCAGTTTGGAGAACATTTATTAAGAGTGTCTACAGACATGGA-GGATTGTCTATCTTG	259		
QY 571	gacatggagagacttaagcggatgtggcgaacaagaagggtcgtactgtcttaagcggatgtcga	630		
Db 260	GCACATGGAGATTATCTGTGGTGGGCCAAGCAAGGTTCTCTCTTCAACGCTGTCTCTCA	319		
QY 631	ccgtgcggggcccatcaaggccaaactccacaagaagacagagcgttggagagacatccacagc	690		
Db 320	CGGTTGCGGCCATCAACGCCAACCTCTCATTAAGAGACGAGCGTGGGAGCAGTTCACTGATG	379		
QY 691	ctgtgataaagtgtgctgtgagcgttcaacccgggaagaagtgcgttctctgtgtgtggggtcat	750		
Db 380	CAGTGTCTCTGCTGCTTAATCAGAACTTGGAAATGGCGTTGTTTCTTGCGTGGGGCTCTTT	439		
QY 751	acgcccaataagaaaggagcgaaccatcgcacagaagaacgctacacatgtcttgcaagctgttc	810		
Db 440	ATGCCCAAGAAGAGGCGCAGTGCCTATGATAGGAACGCCACCACTGTACTACAGACGGCTC	499		
QY 811	atccatctccttctgtcgtcctcatcgttgggttcccttctgtgttgtaagcaactctccaaagctc	870		
Db 500	ATCCCTCCCTTTGTCAGCTGTATAGAGCGGTTCTTTGGATGTAGACACTTTTCAAAAGACCA	559		
QY 871	acgggctgtcttaaacatctctggagcggagcctataaaztgaagaacctcaaa	923		
Db 560	ATGAGCTCTCTCAGAGTCTTGCAAGAAGCCCAATTGACTGGAAGAGCTGTGA	612		
RESULT 6				
AUI26319				
DOCS	AUI26319	609 bp	mRNA	linear EST 23-Oct-2000
DEFINITION	AUI26319 NT2RP1 Homo sapiens cDNA clone NT2RP1000586 5', mRNA			
ACCESSION	AUI26319			
VERSION	AUI26319.1	GI:10951035		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryote: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;			
AUTHORS	Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.			
TITLE	1 (bases 1 to 609)			
COMMENT	Ota,T., Nishikawa,T., Suzuki,Y., Kawai,Y., Ishii,S., Saito,K.,			
	Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.,			
	HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Kawai,Y.,			
	Ishii,S., Saito,K., Nakamura,Y., Nagai,T., Sugano,S., Isogai,T.)			
	Unpublished (2000)			
	Contact: Takao Isogai			

Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES

Location/Qualifiers

source
 1. 609
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RP1000586"
 /clone_1ib="NT2RP1"
 /cell_type="teratocarcinoma"
 /note="Vector: pUC19FL3; mRNA from NT2 neuronal precursor
 cells after 48-hours retinoic acid (RA) induction"
 BASE COUNT 152 a 142 c 144 g 163 t 8 others
 ORIGIN

Query Match 24.0%; Score 307.6; DB 9; Length 609;
 Best Local Similarity 69.8%; Pred. No. 4.3e-65;
 Matches 423; Conservative 0; Mismatches 182; Indels 1; Gaps 1;

OY 310 acttcaacaattgattcttcttgatgatgagagagccgacacgcttaccac 369
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 4 ATTTTATCAANCTAATGAGATTGTTTCAGAGAGAGAGAGATACAGCTTTATCCAC 63
 OY 370 cggctgacagtgatgacagtcagacagagatgtgtgacatcaagatgtgaagtaga 429
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 64 CCCACACAGAGTCTTACCTGAGCCAGATGTGTGACATTAAGATGTGAAGTTTGCA 123
 OY 430 ttctagcagagacccctaccaggtgcccaacacagacatgtactgtttcagttgc 489
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 124 TCTCGGAGAGATCCATATCATGAGACCTATATCAAGCTCAGGCTCTGCTTTAGTTC 183
 OY 490 aaaaagcagttcccccctccacagctcgtgaaacatacaagaattgtgacgaca 549
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 184 AAGGCTGTTCCTCCCTCCGCGCCAGTTTGAGAGAACTTTATAAGATTTGCTACAGACA 243
 OY 550 ttgatgctcaagacalcctcgtgacatgagatctaagcggatgagcaaaacaaagggtgc 609
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 244 TAGAGATTTTGTTCATCCTGCGCATGAGANAATTATCTGGGTGGCCAGCAAGGTGTC 303
 OY 610 tgcgtcctaagcgggtgtcgaacgtgctggccctaccagcccaactcccaagaagacagag 669
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 304 TCTTCTCAACGCTGCTCTCAGGTTGTGCCCATCAAGCCAACTTCATTAAGAGCGGAG 363
 OY 670 gctggagacatcaaccagcgtgtga tcaagtgctgagcgcgtcaaccgggaagagtcg 729
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 364 GCTGGAGAGATTTCAGATGCAAGTTGTCTGTGCTAAATCAGAACTCCAGATGGCTTG 423
 OY 730 ttctcctgtgtgggtcctacagccataaagaaggagcagacatgcagagaagaagtc 789
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 424 TTTTCTTGTGCTGGGCTCTTATGCTCAGAAAAAGGCGACAGTCATGATGAGAGCGGC 483
 OY 790 acaatgcttgcaagcgtgtcctacatccatttctgtctgctacgtggttctcttggt 849
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 484 ACCATGCTTACAGACGCTCATCTCCCTCTTGTGAGTGAAGAGGCTCTTTTGAT 543
 OY 850 gtaagcaactctcaagcgtgacagcgtgtgaaactatctggaagcggagcctaact 909
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 544 GTNNACACTTTTCAAAACCAATGANTGCTG -CAGAACTCTGGCAAGAACCCATTGACT 602
 OY 910 gggagag 915
 |||||
 Db 603 GGAANG 608

RESULT 7
 LOCUS BM449708 881 bp mRNA linear EST 05-FEB-2002
 DEFINITION AGENCOURT_6392952 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528414
 5', mRNA sequence.
 ACCESSION BM449708
 VERSION BM449708.1 GI:18498748
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS NIH-MGC
 TITLE NIH-MGC http://mhc.nci.nih.gov/
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-f@mail.nih.gov
 Tissue Procurement: ATCC/DICP/DMP
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM12205 row: d column: 15
 High quality sequence stop: 623.

FEATURES
 source
 1. 881
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5528414"
 /clone_1ib="NIH_MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

BASE COUNT 219 a 209 c 220 g 231 t 2 others
 ORIGIN

Query Match 23.3%; Score 298.4; DB 10; Length 881;
 Best Local Similarity 71.0%; Pred. No. 8.6e-63;
 Matches 395; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

OY 368 accggtgatcaagtgtacagttctgacagagatgtgtgacatcaagatgtgaagtagt 427
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 4 ACGGCTGGCCCAAGTCTTCACTGACCCAGATGTGTGACATTAAGATGGAAGTTGT 63
 OY 428 gatctagcgaagaccctaccaggtgcccaacacagacatgtactgtttcagtg 487
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 64 CATCTTGGAGACGATCATATCATATGACCAATCAAGCTCAGCGCTCTTATGCT 123
 OY 488 gcaaaagccagttcccccctccacagctcgtgtgaacatatacaagaatgtgtaccga 547
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 124 TCAAGAGCTGTTCCTCCCTCCGCGCCAGTTTGAGAGAACTTTATAAGATGCTTACAGA 183
 OY 548 catgatggtcttaagatcctgacatgagatctaagcggatgagcaaaacaaagggt 607
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 184 CATAGAGATTTTGTTCATCTGCGCATGAGATTTATCTGGGTGGCCAGCAAGGTGT 243
 OY 608 gctgctcttaacgcggtgtcgtgacggtgctgagccacacagcgaaccccaagaagacag 667
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 244 TCTCTTCTCAAGCGCTGCTCAGAGTTGCTGCGCCATCAAGCAACTCTCATAGAGCG 303
 OY 668 aggtcggagacatcaaccagcgtgtgacagatgtgagcgtgacacgggaagagtc 727
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 304 AGGCTGGAGACGATTCATGATGAGTTGTCTCTGCTAAATCAGAACTGGAATGGCT 363
 OY 728 cgtttcctgtgtgggtccatagcgcataaagaaggagcagacatgcagcggaaagc 787
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db	364	TCCTTTTCTTGCCTCGGGGCGCTCTTATGCTCCAGAAAGGCGCATGGCCATTGATAGCAAGCG	423
QY	788	taacatatgtcttgcgaagctgtgtcatccatctcttctgtctgtcatcgttgggttccttgg	847
Db	424	GCACCATGATGACTACAGACAGCGCTCATCCCTCCCTTGTGTACGTATAGAGGGCTTCTTTGG	483
QY	848	ttgttaagcattcttccaaaggtcaagggtctgtctgaacctatcttgagcagagccataaa	907
Db	484	ATGTAGACACTTTTCAAAGAACCAATGAGCTGCTGTCGAGAAGTCTTGCAAGAACCCATTGA	543
QY	908	ctggaagagcactctaa 923	
Db	544	CTGGAAGAGAGCTGTGA 559	
RESULT	8		
LOCUS	BI066406		
DEFINITION	BI066406 622 bp mRNA linear EST 15-JUN-2001 pgfln.pk008.a20 normalized chicken fat cDNA library Gallus gallus cDNA clone pgfln.pk008.a20 5' similar to emb CAA61578.1 (X89398) uracil-DNA-glycosylase, UNG2 [Homo sapiens] emb CAA70211.1 (Y09008)) uracil-DNA glycosylase [Homo sapiens], mRNA sequence.		
ACCESSION	BI066406		
VERSION	BI066406.1	GI:14473928	
KEYWORDS	EST.		
SOURCE	chicken.		
ORGANISM	Gallus gallus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. 1 (bases 1 to 622) Cogburn,L.A., Morgan,R.W. and Burnside,J. Chicken ESTs from fat Unpublished (2001) Contact: Larry A. Cogburn University of Delaware Townsend Hall, Newark, DE 19717, USA Tel: 302-831-1335 Fax: 302-831-2822 Email: cogburnudel.edu, www.chickest.udel.edu.		
JOURNAL			
COMMENT			
FEATURES			
source	1..622 location/Qualifiers 1..622 /organism="Gallus gallus" /db_xref="taxon:9031" /clone="pgfln.pk008.a20" /clone_lib="normalized chicken fat cDNA library" /sex="Male and Female" /tissue_type="fat" /lab_host="E.coli EMDH10B" /note="Vector: pSPORI1"		
BASE COUNT	137 a	165 c	194 g 119 t 7 others
ORIGIN			
Query Match	23.0%;	Score 295.4;	DB:10; Length 622;
Best Local Similarity	70.0%;	Pred. No. 4,3e-62;	
Matches 392;	Conservative 0;	Mismatches 168;	Indels 0; Gaps 0;
QY	186	gagcagctgtaagaatgagccaaataaagaagcagcgttgacaagatagagcaaa	245
Db	63	GAGCAGCTGAGGAGCGGATCGCAGGAACAAAGAGTAGCGCTGCGGAGGCTGAACGAGCGC	122
QY	246	gcaaaagccttgaaagtttggagagaacttggagagaagagagcgtgtctgcagaatttgaagaag	305
Db	123	AGCGTTCTCCGGGGCTTGGGGAGAGCGTGGAGGGCGCGCTGGCGCGCAAGTTTCAACCAAG	182
QY	306	ccatattccaaacaatatgtatcttcttgaactatgagagaagacgcgtcacacgcgtcac	365
Db	183	CCCTACTTCGTGAGAGCTATGGCAATTCGTGTGACCGAGAGAGAGAGAGGTACACGCTGTAC	242
QY	366	ccaccgagctatcaagtgatcaagtcagcagagatggtgtgacattcaagaatgtgaagta	425
Db	243	CCGGCCCCGGAGAGGCTTTCACCTTGACACAGATGTGTCGACATCAGGAGGATGTAAGATTT	302

QY	426	gfgatctctagagccagagacccttaaccacggtcccaaccagaacatgactctgttccagt	485
Db	303	GTAAATCTTG6GCAAAATATCTTAATCAATGAGACCTTAATCAAGATCAATAGGCTGTGTTTACG	362
QY	486	gtgcacaaagccagttccctccctcccccagttctcgtgaacatatacaaaagatgtgtacc	545
Db	363	GTCCAGAAACCTGTTTCCGGCTCCGCCAATTTTGAAACATTTACAAAGACGTGTGACT	422
QY	546	gacattgatgtccttaagaacatcctgtgacatgtgagatctaagcgatgtggcacaagaagg	605
Db	423	GATATTGAAGACTTCACCATCTGTGTCACGGGAGATCTACTGTGGCTGGGCGCAAGCAGGCT	482
QY	606	gtgcgtgtccttaaacgaggtgtgtgacctgtgacgtgtggcccatcagggcaaccccaagaagc	665
Db	483	GGCTCTGTCTCAACCCCTCTCTCAAGGTCGAGCCACAGGCCACGTCCACAAAGAGAG	542
QY	666	agaagctgtggaagacctcacacgcagcgtgtgataaagtgtgctgagcgtcaacgcggagaaga	725
Db	543	AAGGCTGGGACACAGTTCACAGATGTGTGTCTCTGCTGCTGAACAAAGACCTGCGNNGN	602
QY	726	gtcgttttcctgtgtgtggg	745
Db	603	NNTGTTCATGCTGCTGNGG	622
RESULT 9			
LOCUS	AJ394089	792 bp	mRNA linear EST 25-JAN-2001
DEFINITION	AJ394089 dkf426 Gallus gallus CDNA clone 17p6r1, mRNA sequence.		
ACCESSION	AJ394089		
VERSION	AJ394089.1 GI:7123633		
KEYWORDS	EST.		
SOURCE	chicken.		
ORGANISM	Gallus gallus		
	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.		
REFERENCE	1 (bases 1 to 792)		
AUTHORS	Abdrakmanov,I., Lodygin,D., Geroth,P., Arakawa,H., Law,A., Plachy,J., Korn,B. and Buerstedde,J.M.		
	A large database of chicken burral ESTs as a resource for the analysis of vertebrate gene function		
	Genome Res. 10 (12), 2062-2069 (2000)		
JOURNAL	20568495		
MEDLINE	Contact: Buerstedde JM		
COMMENT	Cellular Immunology		
	Heinrich-Pette-Institute		
	Martinistr. 52, 20251 Hamburg, Germany		
	Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.		
FEATURES	Location/Qualifiers		
source	1..792		
	/organism="Gallus gallus"		
	/strain="CB"		
	/db_xref="taxon:9031"		
	/clone="17p6r1"		
	/clone_lib="dkf426"		
	/tissue_type="Bursa of Fabricius"		
BASE COUNT	170 a 225 c 247 g 148 t	2 others	
ORIGIN			
Query Match	22.9%	Score 293.4;	DB 9; Length 792;
Best Local Similarity	70.1%	Pred. No. 1.4e-61;	
Matches 407; Conservative	0;	Mismatches 173;	Indels 1; Gaps 1;
QY	186	gagcagctgtggaagaatgtgccaataataagaagcagcgttcagacaagattagagcaaa	245
Db	206	GAGCAGCTGTGGAAGCGATGCGCAGGAACAAGAGTACGCGCTCGGAGGCTGACACAGCGC	265
QY	246	gcaacgcctgtaggttttcgagagacattgtgagaagagagcgtgtgctgcagagtttgaag	305
Db	266	AGCGTCTCCGCGGCTGTGGGAGAGCTGTGGCGCGGCGGACGTGGCGGACAGTTTCAACAG	325

FEATURES

Location/Qualifiers

1. 863

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS06004YL19"

/clone_lib="LTI_FLO11_BCI"

/sex="male"

/tissue_type="B cells from Burkitt lymphoma"

/lab_host="DH10B"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 188 a 242 c 250 g 182 t 1 others

ORIGIN

Query Match 21.8%; Score 280.2; DB 9; Length 863;

Best Local Similarity 66.2%; Pred. No. 2.6e-58;

Matches 419; Conservative 0; Mismatches 213; Indels 1; Gaps 1;

159 aagacgtcatcgccacagcttcagtgagcagctgagaagaatgagccaaataagaata 218

218 ACCGCGGCTCCTCCGCGCTGAGTGGCCGAGCAGTTGGACGGATCCAGAGAAAGGCC 277

219 gcagcgcttgacaaagattagagcaaaagcaagcgcttcaggttcggagagacttgaga 278

278 GCGGCGCTGCTCAGACATCGCGCCGCCAGACGTGCGCTTGAGAGAGCTGGAAG 337

279 agagagctgctcagaggtttgaaaagccatctcaacaatgagtccttgtagct 338

338 AAGCACCCTCAGCGGGGAGTGGGAACCGTATTATTATCAAGCTAATGGATTTGTTGCA 397

339 gatgagagagcgctcaccacgctctaccacgcgctgatcaagtgatcaggttcagagag 398

398 GAAGAAGAAAGCATATGATGTTTATCCACCCACACCAAGTCTCAGCTGAGCCAG 457

399 atgtgtgacattcaagaatgtgaagtatgattcttaggccaagagaccttaccacggtccc 458

458 ATGTGTGACATTAAGAGTGTGAAGGTTGTATCTGGAGACGATTCATATCATGAGACT 517

459 aaccaagcatatgactcttctcaagtgtgcaaaagccagttcccccctccacagctc 518

518 AATCAAGCTCAGCGGCTCTGCTTATAGTGTCAAAAGGCTTCCGCTCCGCCAGTTTG 577

519 gtgaacatatacaagaatgtgttaccagacattgagcttcaagcaltccctgagacatgga 578

578 GAGAACAATTATTAAGAGTGTCTACAGACATAGAGATTTTGTTCATTCCTGGCCATGGA 637

579 gatctaagcggaatgggcaaaaggggtgctgctcttaacgcggtgctgagccgtcgga 638

638 GATTATATCTGGGGGCGCAAGAGTGTCTCTCTCAACGCTGCTCCACAGCTTGTG 697

639 gccatcatgagccaactcccaagaagagagctgagagaccttaccagagcgctgagtc 698

698 GCCCATCAACCAACTCTCTTAAGAGCGAGGCTGGAGACAGTTCACTGATGCACTTTGTG 757

699 aagtgagctgagcgtlcaaccgggaagagtgctttctcgtgtgtgaggtctatacggccat 758

758 TCTTGCTAATATCAGAACTGGAATGGCTTGTCTTCTGCTGGGGCTCTTATGCTCAG 817

759 aagaagagcagacatcgacaggaacgtcac 791

818 AAGAAGGG-CAGTGCATTGTATGGAAGCGGCAC 849

LOCUS B1651753 788 bp mRNA linear EST 12-SEP-2001

DEFINITION 603298541F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5339027 5', mRNA sequence.

ACCESSION B1651753

VERSION B1651753.1 GI:15565989

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 788)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LHAM1861 row: a column: 12

High quality sequence stop: 788.

FEATURES

source

1. 788

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:5339027"

/clone_lib="NIH_CGAP_Mam3"

/tissue_type="tumor, gross tissue"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; NotI; Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert 2 kb. Library constructed by Life Technologies, catalog #12017-018. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999). Note: this is a NC1_CGAP Library."

BASE COUNT 174 a 236 c 206 g 172 t

ORIGIN

Query Match 21.4%; Score 274.4; DB 10; Length 788;

Best Local Similarity 70.2%; Pred. No. 6.9e-57;

Matches 368; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

403 gtgacattcaagaatgtgaagttagtgaattctaggccaagagaccttaccaggtcccaac 462

13 GCGCATCCGAGATGTAAGGTTGTCATTTCTGGAGACGATCCCTATCAGGAGACTATATC 72

463 aagacatgagactctgttctcagtgatgcaaaaagcagttcccccctcccccagctcgtga 522

73 AAGCTCAGGGGCTCTGCTTATAGTGTCCAAAGACCAAGTTCCACCCGCCAGTTTGGAAA 132

523 acatatacaagaatgtgttaccagacattgaggttccaagcaltccctgagacatgagatc 582

133 ACATTTTAAAGAGCTGTCTACAGACATGATGTTTGTTCATTCCTGCGCATGGGATT 192

583 taagcgatgtggcaaaacaaaggggtgctgctcttaacgcggtgtcgtacgttgagggccc 642

193 TGTGAGGGTGGGCGCCGCAAGAGTGTCTCTCTCCCAACGCGGCTCTCACTGCTCGGCCCC 252

643 atcagcgaactccccaagaagagagctgagagaccttaccagagcgctgtgataagt 702

253 ACCAAGCCAATTCCTCAATAGGAGAGGGGCTGGAGCAGTTCACGAGACGGCTGCTGCT 312

703 ggtcagcgtcaaccgggaagagtgctttctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 762

313 GCGTGAATCAGAACTCAGTGGCTGCTGCTTCCTGCTGCGGAGCTCTTATGCTCAGAGA 372

763 aggaagcgaccatcgacaggaacgtcacatgtcttgcaagctgttcatcatcattcctt 822

RESULT 12

B1651753

Db	373	AGGCGAAGTCAATTGACAGAGAAGCGTACACATGTTCTCTCAGACACACTCTACCCCTCTCCGCG	432
Qy	823	tgctgtctcatccgtygggttctcttggtytgaagcacttctccaaggctcaaggctgtctgta	882
Db	433	TGTCGGTCTACAGAGAGCGTTCCTGGATCAGACAACTTTTCTAAAGCCATGAACACTGCCTCC	492
Qy	883	aactatctggagcagagccttaaaccttgagagacacttaactc	926
Db	493	AGAAGTCTGGCAGAGAGCCCATCACTGGAAGGACTGTGACCC	536
RESULT	13		
LOCUS	B1940552		
DEFINITION	dbj11f06.y1 Wellcome CRC psk egg Xenopus laevis cDNA clone	574 bp	linear
ACCESSION	B1940552		
VERSION	B1940552.1		
KEYWORDS	EST		
SOURCE	African clawed frog.		
ORGANISM	Xenopus laevis		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Anphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.		
AUTHORS	1 (bases 1 to 574) Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Miller,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.		
TITLE	Washu Xenopus EST project, 1999		
JOURNAL	Unpublished (1999)		
COMMENT	Other_ESTs: dbj11f06.x1 Contact: Sandy Clifton, Ph.D. Washu Xenopus EST project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Library constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Xenopus clones from this library are available through the I.M.R.G.-E. Consortium/LLNL at: info@image.llnl.gov Seq primer: -40RP from Gldco High quality sequence stop: 454. Location/Qualifiers 1. 574 /organism="Xenopus laevis" /db_xref="taxon:8355" /clone="IMAGE:3378467" /clone_id="Wellcome CRC psk egg" /tissue="egg" /lab_host="DH10B (phage-resistant)" /note="Vector: pBluescript SK-; Site.1: NotI; Site.2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Library was constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute). Note: This is a Xenopus Gene Collection (XGC) library."		
FEATURES			
SOURCE			
BASE COUNT	170 a	132 c	136 g
ORIGIN			136 t
Query Match	21.2%	Score 272.6;	DB 10; Length 574;
Best Local Similarity	67.5%;	Pred. No. 1.8e-56;	
Matches 383;	Conservative 0;	Mismatches 184;	Indels 0;
Gaps 0;			
273	tgaggagaagagcgtggcgcgcagagttgaaaaagccatactccaacaatgatgtccctt	332	
8	TGGAACAAGGAGAGCTCGTGGATTTTCACAAACCCACTTGTTCACACTCAATTC	67	

QY	333	gtacgtgtatgagagagacgcctcaacacccgtccaccacacccggtctgtatcaagtgtaagttcg	392
Db	68	ATTGCAGAGAGAGAGAAAGAAATGCACACTTTATATCCACCACAGAGAGATATCACTGG	127
QY	393	acagagatgtgtgacatccaagatgtgaaagtaigtatcttcaggaccagacccttacac	452
Db	128	ACACAAATGGTTGACATCAAGATGTAAAGGTGTTATTTTAAAGGCCAAGATCTTATCAT	187
QY	453	gttcccaaccagaacacatgagactctgtttcagttgtgtgcaaaagccagttcccccctcc	512
Db	188	GGACCAATATCAAGCTCATAGTGCTTTGGCTTTAGTGTCAAGAAAGCCAGTCCACCTCCACC	247
QY	513	agtcctgtgaacataatacaagaattgtgtaccagacatgtatgtcttcaagcatccctg	572
Db	248	AGCTTAGTGAATATATACAAAGAACTTACAAMCCGATATCGAAGATTTTTCAGCCCAAGT	307
QY	573	catgagaatctaaacggaatggtgcaaaacaaagggatgtctgtctgtctaacgcggtgtgacc	632
Db	308	CATGGGGATCTTCACAGGGCTGGGCCAAACAGAGAGTCTCTCCCTCAATGCTGTGTGACA	367
QY	633	gtgcggagcccatcgaagcccaactcccaacaaaggacagaggcttgagagacttaaccagact	692
Db	368	GTGCAGACTCATATATGCAAACTCTTACACAGCACTGTGATGGGAACAATTCCACAGATGTT	427
QY	693	gtatcaaatgtgtgtgagcgtcaacccgggaagagtcgtttcttcgtttgttggggtcatcac	752
Db	438	GTGGTATCTCGTGGCTTAATAGAACATGATGGGTGGTTTCATGCTGTGGGAGACTTAT	487
QY	753	gcccaataagaaggagcgacacatgcagcaggaagcagtcaccatgtcttcgaagctgttat	812
Db	488	GCCCAAGAAAGGAGCAGCAATATTTACAGAAAGGACACAGTCTTCAAAACCGTGCAC	547
QY	813	ccatctccttctgtctgcctcatcgttgg	839
Db	548	CCTTCTCCTCTCTCTGATGCTCACAGAGCG	574
RESULT 14			
Bi657579		768 bp	linear
LOCUS			EST 12-SEP-2001
DEFINITION	60328153F01 NIH_CGAP_Mam4 Mus musculus cdna clone IMAGE:5325848 5',		
ACCESSION	Bi657579		
VERSION	Bi657579.1		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota: Metazoa; Chordata: Craniata: Vertebrata: Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 768)		
JOURNAL	NIH-MGC http://mgc.ncl.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgaps-remail.nih.gov		
	Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch		
	Ph.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLN at:		
	http://image.lnl.nih.gov		
	Plate: L1AM11826		
	row: 1		
	column: 09		
	High quality sequence stop: 764.		
FEATURES			
source	location/Qualifiers		
	1..768		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:5325848"		
	/clone_id="NIH_CGAP_Mam4"		
	/tissue_type="tumor, gross tissue"		
	/lab_host="DH10B"		

/note="Organ: mammary; Vector: PCMV-SPORT6; site_1: NotI; site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert 2.5 kb. Library constructed by Life Technologies, catalog # 12018-016. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP Library."

BASE COUNT
ORIGIN

Query Match 20.8%; Score 267.2; DB 10; Length 768;
Best Local Similarity 66.7%; Pred. No. 4.1e-55;
Matches 411; Conservative 0; Mismatches 203; Indels 2; Gaps 2;

QY 150 gtgagcaaaagacgcatcgccacagcttcctagtgagagctggaagaatgagccaa 209
DB 154 GTGAGCAGACGACGACGAGGCTCGCGCTCAGCGCCGACGCTCGCCATCCAGAGG 213
QY 210 aataaanaagcagcgcttgagcaaatagagcaaaagcaagcgcctgcaggttcgagag 269
DB 214 AACAGGCGCGCGCGCTCGCTCAGGCTCGCCCGCAGCTGCTCGCGGCTTCGGGAG 273
QY 270 acttgagaagagagctgctgcagagtttgaaagccactcaacaatgatgtcc 329
DB 274 AGCTGGAGCAGCAGCTGTGTCGGGGAGTTGGGAGCCGCTACTTCCTCAAGCTAATGGA 333
QY 330 ttgtgagctgagagagagcgctcaacacgcttaccacagcgctgagcaagtgaag 389
DB 334 TTTGTGCTCAGAAAGAGAACACACACAGGCTATCTCCGCCCGGAGGAGTTCACA 393
QY 390 tcgacagagaatgtgtacatcaagatgtgaagtagtgaattcaggccagacccttac 449
DB 394 TGAGCCAGCATGTGCGACATCCGAGATGTGAAGTTGTATCTCGGAGACAGATCCTAT 453
QY 450 caaggtcccaacgaagcaactggaactgttcaagtgtgcaaaagccaggtccccc 509
DB 454 CAGGACCTAATACGTACGGGCTCTGCTGCTCAAGTGTCCAAAGACCAAGTTCACACCCCG 513
QY 510 cccagctctgtagacatatacaagaattgtgacgacatgtagcttcaagcaccct 569
DB 514 CCCAGTTTGGAAAACATTTTAAAGAGCTGTCTACAGACATCGATGG-TTTGTTTCACTCT 572
QY 570 ggaatgagatctaaagcggaatgggcaaaaggggtgctgtcgtttaacgcggtgtc 629
DB 573 GGGCAATGGGATTGTGACGGGTGGGCCCGACAAAGGTGTCTCTCCCTCAAGCGCTCTCTC 632
QY 630 accgttcggggccatcaagcaactccacaagaagagagcggtggaagactccacgac 689
DB 633 ACTGTCCGGCCCAACCAAGCAATTCCTCATAAAGAGAGGGGCTGGGAGCAGTTTACCGGAC 692
QY 690 gctgtatcaatgagctgagcgtlcaacccgggaagagagcttcttcctgtgtg-99gctc 748
DB 693 GCGGTGCTGTCCTGGCTGATCAGAACCTAGTGGCTCTCTTCTCTCTCTGTGGGCTC 752
QY 749 ataagcccataagaag 764
DB 753 TTATGCTCAGAGAGCG 768

RESULT 15

LOCUS BG866388 614 bp mRNA linear EST 29-MAY-2001
DEFINITION 602185307F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4911553 5',
ACCESSION BG866388
VERSION BG866388
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 614)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabos-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHM10813 row: n column: 02
High quality sequence stop: 614.
Location/Qualifiers

FEATURES

source 1..614
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4911553"
/clone_1db="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Salivary gland; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 137 a 169 c 168 g 140 t

Query Match 20.5%; Score 262.4; DB 10; Length 614;
Best Local Similarity 71.5%; Pred. No. 5.8e-54;
Matches 372; Conservative 0; Mismatches 146; Indels 2; Gaps 2;

QY 331 ttgtagctgtagagagcgctcaacacgcttaccacagcgctgtagcaagtgtacggt 390
DB 97 TTGTGCTGAGAAAGAGAACACACAGGCTATCTCCGCCCGGACAGGTTCACAT 136
QY 391 cgacagagaatgtgacatcaagaatgtgaagtagtgaattcaggccaggaacctacc 450
DB 157 GGACCCAGATGTGGACATCCGAGATGTGAAGTGTTCATTTGGGACAGATCCTATC 216
QY 451 acggtcccaacgaagcaactggaactgttcaagtgtgcaaaagccaggtccccc 510
DB 217 ACAGACCTAATCAAGCTACGGGCTCTGCTCAAGTGTCCAAAGACCAAGTTCACACCCCG 276
QY 511 ccaagctctgtagacatatacaagaattgtgacgacatgtagcttcaagcaccct 570
DB 277 CCAAGT-TGAAAACATTTTAAAGAGCTGTCTACAGACATCGATGG-TTTGTTTCACTCTG 334
QY 571 gacatgagatctaaagcggaatgggcaaaaggggtgctgtcgtttaacgcggtgtcga 630
DB 335 GCATGGGAGATTGTGACGCTGGGCCCGCACAAGGTGCTCTCTCCCTCAAGCGGCTCTCA 394
QY 631 ccgttcggggccatcaagccaactccacaagaagagagctggggaaccttcaacgagc 690
DB 395 CTGTCCCGCCCAACCAAGCAATTCCTATGAAGAGAGGGGCTGGGACAGTTCACCGGACG 454
QY 691 ctgtgataaggtgtagcgtlcaacccgggaagagagcttcttcctgtgtg-99gctc 750
DB 455 GGGTGTGCTCGCTGATCAACAGACCTGAGTGGCTCTGCTCTCTCTCTGAGGCTCTT 514
QY 751 acgccaataagagagagcgaccatcagcaagaagaacgctcaacatgtcttcaagctgtc 810
DB 515 ATGCTCAGAGAGAGGAGAGGTGATTCATACAGAGAGGCTACACATGTTCTCGACAGACACTC 574
QY 811 atccatctccttgcctgcacatcggtgtgtcctgtgtg 850
DB 575 ACCCTCCCGCTGTGCTGATACAGAGGTTCTCTGGGATG 614

Search completed: August 22, 2002, 04:25:26
Job time: 8136 sec

APPLICANT: Ohki, Misao
APPLICANT: Kikuchi, Kimiko
APPLICANT: Miyoshi, Hiroyuki
APPLICANT: Kozu, Tomoko
TITLE OF INVENTION: No. 5580727e1 DNAs, Polypeptides Encoded Thereby
TITLE OF INVENTION: and Methods for Detecting the DNAs and Polypeptides
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 810 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/244,189
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 760-183P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 752 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-244-189-2

Query Match 5.6%; Score 89; DB 1; Length 752;
Best Local Similarity 21.8%; Pred. No. 0.67;
Matches 69; Conservative 36; Mismatches 82; Indels 130; Gaps 20;
QY 18 LPLGILLPQTLCSKLMKTPPKLRSSN---VEQRTSSPOLSVQLEEMANN-KKAALDK 72
DB 337 LP-LLORELLHCARLAKONPAQYLAQHEQLLDASTSPVDSSELLDVNENGRKRPDR 395
QY 73 IRAKATGAGFETWRRELAFFEKPYFKQLMSFVADERSRTVVPRAQVVSSTEMCIDIQ 132
DB 396 TKEN---GF-----DREPLHSEHP-----SKRPTIS 419
QY 133 DVKVVILGODEYHGPNOAHGLCF---SVQKPVPPP-----SLVNIYKELCTDIGF 181
DB 420 P-----GQ--RYSFN--NGLSTQPNGLPHPTPPRPGHRIIDMAIAHHYR-----DSY 463
QY 182 KHPHGDLG-----WAKGV---LLNAV-----L 204
DB 464 RHPSHRDLRDRNPRMGLHGTROEEMIDHRLDRMAEEMKMLDHLNCLIMDVEKTRSL 523
QY 205 TV--RAHQANSHKDRGW-ETTTDAVIAKLSVNRGVVFLMGSTAHKKGATITDKRHNV 261
DB 524 TVLRRCGEADEELNYWIRIRSDA-----BDL-----KKGCGSSSSSHRQ 564
QY 262 QAVHSP--LSAHGFL 276
DB 565 SPVNPDPVALDAHREFL 581
RESULT 3
US-08-826-267-2
; Sequence 2, Application US/08826267
; Patent No. 5994070

GENERAL INFORMATION:
APPLICANT: Streuli, Michel
TITLE OF INVENTION: NO. 5994070e1 TRIO Molecules and Uses Related Thereto
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,267
FILING DATE: 1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,214
FILING DATE: 27 MARCH (1996)
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2860 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-826-267-2

Query Match 5.1%; Score 81.5; DB 2; Length 2860;
Best Local Similarity 19.9%; Pred. No. 33;
Matches 59; Conservative 41; Mismatches 95; Indels 101; Gaps 13;
QY 12 ISSNRVLPGLLPQTLCSKLMKTPPKLRSSNVE---QKRTSSPOLSVQLEEMAKNK-- 66
DB 684 ISSNK-----TPHNSSTNHETVLDQLEDAOSQWEELEFQERKIKLE 724
QY 67 -----KALDKIRAKATPAGFETWRRELAFFEKPYFKOLMSF-----VADERS 111
DB 725 LFLHVRIFERDAIDILISL-----ESWNDELS-----QWNPDTEDLTIAEQL 769
QY 112 RHTVVPRAQVVSSTEMCIDIQDVKVVILGOD--PYHGPNOAHGLCFSVQKPVPPPSLVN 169
DB 770 QH-----HADKALTMNLT-----TFDVIHQGDLLQVNVNVAOSGVALLCDRVDATRVQD 821
QY 170 IY-----KELCTDIGEFKHPGH-----GLSGMAKOGVLLNA-VLTVRAN 209
DB 822 LLEFLHEKQOELDLDAEORHKLLEQCYOLRHLDQAEVKQVLTWIRNGEEMLNAGLITASSL 881
QY 210 QANSHKDRGWETTTDAVIAKLSVNRGVVFLMGSTAHKKGATITDKRHNVIAVNH 265
DB 882 QEADLOREHNOFOHAIEK-----THOSALOYOOAKEMALQANH 920
RESULT 4
US-08-631-607-6
; Sequence 6, Application US/08631607
; Patent No. 5767252
; GENERAL INFORMATION:
; APPLICANT: Morley et al, Paul
; TITLE OF INVENTION: NOVEL NEURONAL CELL GROWTH FACTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/631,607
FILING DATE: 08-APR-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/086001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 426 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-631-607-6

Query Match 5.18; Score 81; DB 1; Length 426;
Best Local Similarity 22.7%; Pred. No. 2.1;
Matches 63; Conservative 37; Mismatches 102; Indels 76; Gaps 15;

QY 29 FSKLMKTPKRLRSNV---EQKTS-----SPQLSVQLERMAKNKAALD--KIR 74
DB 81 FSAIRELTSKLARCEGLMAGAESSKDTMGDLPRDPSPRVQLSRSLQVLDRLSELELR 140
QY 75 AKAT----PAGFGEYWRRELAAFEKPYFKOLMSFVA---DERS-RHTVYPADQVVSST 126
DB 141 TMSNTGTPSPFRVLDQRL-GELE---KQLLKVALEBEKSLINNETSAHQOKTSNT 195
QY 127 EMCIDIQVKVVIIGDPPYHGN-----QAHGLCFSVQKVPVPPSLVNIYK-ELCTD 177
DB 196 LNALQRTTELEGRNSAFKSPDAFKVSLPRTTYLYGKIKKTLR-----ELYSFTICLM 249
QY 178 IDGFKHGHGDLGSMWAG-----VLLN---AVLYTRAHQANSH----- 214
DB 250 LRSSASPGIGTPFSYAVPGGANETVLIEWGNPIELLINDKVAQLPFLVSDGKWHHICIT 309
QY 215 ---KDRGMEFTTDA-----VIKMLSVNREGVFFLL 241
DB 310 WTRDGLMEAFQDEGKLGCTGENTLAPMHPKISGVPLLI 347

RESULT 5
US-09-074-579-3
; Sequence 3, Application US/09074579
; Patent No. 6001596
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,579
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cerrone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0505 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 946 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENEBANK
CLONE: g133985
US-09-074-579-3

Query Match 5.08; Score 80; DB 3; Length 946;
Best Local Similarity 19.18%; Pred. No. 9.1;
Matches 71; Conservative 42; Mismatches 113; Indels 146; Gaps 17;

QY 34 KITPKRLR--SSNVEOKTSSPOLSYEQLERM-----AKNKAALDKIRAKAPAGGGE 84
DB 532 KFDPAKLDQISVITATSAHQVLETLAQMDLDQDFLSKDKHADPPDTR-----K 582
QY 85 TW-----RRELA--AEFEKPYFKOLMSFVADERSRHTVYPPADQVVSSTEMCIDIQ 132
DB 583 LMAVLTINQLAERSLAPTAAKRRTIRSILOMSID---HHIVYPLSLVIENAGDER 638
QY 133 DVKVVIIIGDPPYHGNQAHGCFSVQKVP-----PP 164
DB 639 ---MLADAPPDPSCCSGALVYGSKVVPDSTPSMANPSPTVYISMLAGSQVLESTPP 693
QY 165 PSLVNI-----YKELCTDID-----GFKHPGHG 187
DB 694 PHVWRVENDPHFITLYLPKRSQNICFNIDSEPKILNLVSPDESIGVYNGQLVGAKKRPNG 753
QY 188 DLGSM-AKQGVLLINAVLYTRAHQANSHKDRGMEFTTDAVTKW-----LSVNR 234
DB 754 KLSYFGKLGFFQSEDIKIEISTETITLSHGSSSTFS---LSWSDTAQVTRQVQISYK 810
QY 235 EGVVFLMGSYAHKKKATIDKRRH---HYLQAVHPSPSLAHNGFLGC-----KHKSK 283
DB 811 EKVVYI-----TLDKEMSFVLLHRYWKKHP---VNVDFLGITYIPPTNKSPK 855
QY 284 ANGLIKLSTGP 295
DB 856 AHGLIGQMGEP 867

RESULT 6
US-09-388-774-3
; Sequence 3, Application US/09388774
; Patent No. 6228991
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.

APPLICANT: Guegler, Karl J.
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/388,774
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/074,579
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cerrone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0505 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 946 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENE BANK
CLONE: g133985
US-09-388-774-3

Query Match 5.0%; Score 80; DB 4; Length 946;
Best Local Similarity 19.1%; Pred. No. 9.1;
Matches 71; Conservative 42; Mismatches 113; Indels 146; Gaps 17;
QY 34 KITPKLR--SSNVEQKTSFQLSVEQLER-----AKKKAALDKIRAKATPAGFGE 84
DB 532 KFDPAKLDIESVTATSAHQVLLETLAQMDLDQDFLSKDHADPPFTR-----K 582
QY 85 TW-----RRELA--AEFEKPYKQMSFVADERSRHTVPRADQVSSSTEMCDIQ 132
DB 583 LMAVLTNQLAERSLAPTAAKRRITRSILOMSLD---HHIVTPLSLVIENAGDER 638
QY 133 DVKAVILGDPYHGNPQAHGCFVSQKRPV-----PP 164
DB 639 -----MLADAPRQDPSCSGALYYGSKVVPSTPSMANPSTPIVISMAGSOYLESTPP 693
QY 165 PSLVNI-----KELCTDID-----GKHPGHG 187
DB 694 PHVVRVENDPHFIILPKSQKNICFNIDSEPGILNLVSDPSGIVVNGQLVGAKKPPNG 753
QY 188 DLSGW-AKQGVLLNNAVITVAHQANSHKDRGWTFTDAVTKW-----LSVNR 234
DB 754 KLSYFEGKLGTFYQSEDIKIEISTETITLSHGSSTFS--LSWSDTAQVTVNQRVOISYK 810
QY 235 EGVVFLMGVSAHKKGATIDDKRH-----HYLAHVHPSPLSAHGRGLGC-----KHPSK 283
DB 811 EKVVTT-----TLDEKMSFVLLHRYWKHP-----VNVDFLGITYPINPKTSFK 855
QY 284 ANGLIKLSTGP 295

DB 856 AHGLIGFMQEP 867
RESULT 7
US-08-631-607-7
Sequence 7, Application US/08631607
Patent No. 5767252
GENERAL INFORMATION:
APPLICANT: Morley et al, Paul
TITLE OF INVENTION: NOVEL NEURONAL CELL GROWTH FACTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/631,607
FILING DATE: 08-APR-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/086001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-631-607-7

Query Match 4.9%; Score 79; DB 1; Length 430;
Best Local Similarity 23.9%; Pred. No. 3.6;
Matches 57; Conservative 25; Mismatches 94; Indels 62; Gaps 12;
QY 56 VEQLERNAKKKAALD---KIRAKATPAGGFTWRREL---AAEFKPYKQMSFYA- 107
DB 123 VEQLSRSLQTLKRLSEHQLRANVSAGLPDGFREVLDQRLGELE---RQLLRVAAE 178
QY 108 --DERS-RHYVYPPADQYSSSTEMCDIQDVKVVILGDPYHGNPQAHGICFSYQKRPV- 163
DB 179 LEDERKSLHNETSAHROKTESTLNAALQRYTELRGNSAFKSPD-----AKVSLPLPTN 233
QY 164 -----PPSLVNIYK-ELCTDIDGFKHGHGDLGMAKKG-----VLLLN 201
DB 234 YLYGKIKKTLPELALFTICMLRBSASBPIGTSPSYANPGQANELLILEMGNPIELLIN 293
QY 202 ---AVLTVRAHQANSH-----KDRGWETFTDA-----VIKLSVNRGVEVL 240
DB 294 DKVAQDLPLFVSDGKWMHNCVTWTRDGMWFAFGDEKLGTEGNLAPMHPINPGVLLI 351
RESULT 8
US-08-568-459A-12
Sequence 12, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelisen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-12

Query Match 4.9%; Score 78; DB 2; Length 2710;
Best Local Similarity 22.8%; Pred. No. 74;
Matches 71; Conservative 34; Mismatches 129; Indels 78; Gaps 15;

29 FSKLMKTPKRLSSNVEKTSPOLSVEQLERMAK-----NKKALDKIRAKATPAG 81
1496 FSKDGSKSPSGL-SROEMWKTNGPEIWKGMICALTKYVDTDNKRRIKNDYSTDKVNQSQ 1554
82 FGETWRRELAEEFEKPYFKOLM-----SFVADERSRHTVYPPADQVYSTEMCDIQDVKY 136
1555 NGNPSLEEFNA---KPPFLMWMIEWGEEPCAEGRKKENIIKDCANELNSIQGCN----- 1605
137 VILGDPYHGPNOAHGLCSFYOKVPVPPPSLVNIYKELCTIDIGF-----KHGEGDLSG 191
1606 -----DAKHRCNOA---CRAVYOEVY-----NKKKEFSQTNNFVLKANYQPDDEPKG 1651
192 WA-KOGVLLNLNAVLTVAROAHNSHK-----DRGWETFLDAVIKMLSVNREGVFLMGSYA 246
1652 YEYKDG-----VQPIQGENYLLQKCDNNKSCMDGNV--LSVSPKEKPF---GKYA 1697
247 HKKGATIDRRKHHVLAQVHPSPLSAHRG-----FLGCKHFSKANGLLK 289
1698 HKYPERKCDYQGHKHPISIPRPPRPVQRPQPEAPVTVVDVCSIVKTLFQDTNNFSDACG-LK 1756
290 LSGTEPIIMRAL 301
1757 YGKTAPSSMKCI 1768

RESULT 9

US-08-487-826B-12
Sequence 12, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitlois, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelisen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-12

Query Match 4.9%; Score 78; DB 2; Length 2710;
Best Local Similarity 22.8%; Pred. No. 74;
Matches 71; Conservative 34; Mismatches 129; Indels 78; Gaps 15;

29 FSKLMKTPKRLSSNVEKTSPOLSVEQLERMAK-----NKKALDKIRAKATPAG 81
1496 FSKDGSKSPSGL-SROEMWKTNGPEIWKGMICALTKYVDTDNKRRIKNDYSTDKVNQSQ 1554
82 FGETWRRELAEEFEKPYFKOLM-----SFVADERSRHTVYPPADQVYSTEMCDIQDVKY 136
1555 NGNPSLEEFNA---KPPFLMWMIEWGEEPCAEGRKKENIIKDCANELNSIQGCN----- 1605
137 VILGDPYHGPNOAHGLCSFYOKVPVPPPSLVNIYKELCTIDIGF-----KHGEGDLSG 191
1606 -----DAKHRCNOA---CRAVYOEVY-----NKKKEFSQTNNFVLKANYQPDDEPKG 1651
192 WA-KOGVLLNLNAVLTVAROAHNSHK-----DRGWETFLDAVIKMLSVNREGVFLMGSYA 246
1652 YEYKDG-----VQPIQGENYLLQKCDNNKSCMDGNV--LSVSPKEKPF---GKYA 1697
247 HKKGATIDRRKHHVLAQVHPSPLSAHRG-----FLGCKHFSKANGLLK 289
1698 HKYPERKCDYQGHKHPISIPRPPRPVQRPQPEAPVTVVDVCSIVKTLFQDTNNFSDACG-LK 1756

OY 290 LSGTEPIMRAL 301
DB 1757 YGKTAPSSMKCI 1768

RESULT 10

US-08-487-826B-14
; Sequence 14, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Knobbé Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-826B-14

Query Match 4.9%; Score 78; DB 2; Length 3060;
Best Local Similarity 22.8%; Pred. No. 88; Mismatches 129; Indels 78; Gaps 15;
Matches 71; Conservative 34;

OY 29 FSKIMITPKLSSNVEQRTSSPOLSVLEQLERMAK-----NKRALDPIRAKATPAG 81
DB 1494 FSKDGSKSPGL-SRQEMWKTNPEIMKGLCALTYVDTDKRKIKINDYSYDKVNOQSG 1552
OY 82 FCGTWRELLAEFEKPYFKOLM-----SFVADERSRHITVPPADQYSSSTEMCDDIDQVY 136
DB 1553 NGNPSLEEFALA---KPOFLMMIEMGECECAEROKKENIITDACEINSTQOCN----- 1603
OY 137 VILGDDPYHGPNQAHGCLCFSVOKRPPPSLVNITKELCTDDIGF-----KHFGHGLDLSG 191
DB 1604 -----DAKHCNQA---CRAVQEVVE-----NKKKEFSQTNNEFLKANVQPODPEYKG 1649
OY 192 WA-KQGVLLNNAVLTVRAHQANSK---DRGWEFTTDAVYIKWLSYNRGEVVELMGSYA 246
DB 1650 YEKYGS-----VQPIQGNELLYLQKCDNNKSCMDGNV--LSVSPKRPKPF---GKVA 1695
OY 247 HKGATIDRRHHVYLAHPSPLSAHRG-----FLGCKHFSKANGLLK 289

DB 1696 HKYPEKDCYQAGKAVPSIPPPPPVOPQPEAPVTWYDVCSIVKTLFKDTNNFSDAG-LK 1754
OY 290 LSGTEPIMRAL 301
DB 1755 YGKTAPSSMKCI 1766

RESULT 11

US-08-064-174-1
; Sequence 1, Application US/08064174
; Patent No. 5618540
; GENERAL INFORMATION:
; APPLICANT: QUDENTIN-MILLET, Marie-Jose
; APPLICANT: LISSOLO, Ling
; TITLE OF INVENTION: SUBUNIT VACCINE FOR NEISSERIA
; TITLE OF INVENTION: MENINGITIDIS INFECTIONS AND CORRESPONDING PURIFIED
; TITLE OF INVENTION: SUBUNITS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/064,174
; FILING DATE: 25-MAY-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 12176
; FILING DATE: 03-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 016100-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/836-6620
; TELEFAX: 703/836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 579 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-064-174-1

Query Match 4.8%; Score 77.5; DB 1; Length 579;
Best Local Similarity 26.6%; Pred. No. 8.1;
Matches 45; Conservative 26; Mismatches 69; Indels 29; Gaps 9;

OY 76 KATPAGGTWRRELLAEFEKPYFKOLMSFVADERSR-----HTVPPADQYSSSTEMCDD 130
DB 80 RGEPNFSF--RDLE--EKRGSSLEIESKWEDEGQSRVGYTNFTYVRSGYVTLNNNID 135
OY 131 IDQVYVILGDDPYHGPNQAHGCLCFSVOKRPPPSLVNITK---ELCTD-IDGFRHPGH 186
DB 136 IKN-NIVLEGPBGY-----LYYKKEPSKLLPSEKITYKGTWYVIDAMKQKPEGL 186
OY 187 GDLSGMAKQGVLL--LNAVLTVRAHQANSKDRGWEFT-----FTDAVIR 228
DB 187 GSAAGDDKSGALSALBEGVLRNQAEASSGHTTDRGMTSEFEVDPSDKTIK 235

RESULT 12

```
US-08-066-167-1
; Sequence 1, Application US/08066167
; Patent No. 5618541
; GENERAL INFORMATION:
; APPLICANT: QUENTIN-MILLET, Marie-Jose
; TITLE OF INVENTION: VACCINE AGAINST NEISSERIA MENINGITIDIS
; TITLE OF INVENTION: INFECTIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 2213-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/066,167
FILING DATE: 02-JUN-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 12177
FILING DATE: 03-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feuery, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 016100-002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 579 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Neisseria meningitidis 2394 subunit Tbp2.
US-08-066-167-1

Query Match          4.8%, Score 77.5; DB 1; Length 579;
Best Local Similarity 26.6%, Pred. No. 8.1;
Matches 45; Conservative 26; Mismatches 69; Indels 29; Gaps 9;

QY      76 KATPGPEFTMRRLAEAFKPKFYQLMSFVADERSR-----HTYYPPADQYYSTSEMKD 130
        :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       80 RGEPSFSFE--RDGL--EKRRSSSELLSKWEDGDSRVVGYNFTYVRSGVYLNNNID 135
QY      131 IQDYKVVLIGDDPYHGPNQAAGLCFSPVKPVPPEPLVNLYK---ELCD-IDGFKNPCH 186
        ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       136 IKN-NIVLFGPDGY-----LyyKGKEPSKELPSEKITYGTMGTDYVDANENKQRFESL 186
QY      187 GDLSGMAGQVY--LLNAVLTIVRAQANSKDRGET-----FTDAVTK 228
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       187 GSAAGGDKSGLSALEEGVLRNQAAEASGGHDFGMTSEFEVDFSCKTIK 235

RESULT 13
; Sequence 1, Application US/08449733
; Patent No. 5928650
; GENERAL INFORMATION:
; APPLICANT: QUENTIN-MILLET, Marie-Jose
; TITLE OF INVENTION: LISOLO, Ling
; TITLE OF INVENTION: SUBUNIT VACCINE FOR NEISSERIA
; TITLE OF INVENTION: MENINGITIDIS INFECTIONS AND CORRESPONDING PURIFIED
; SUBUNTS
```

```

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,733
FILING DATE: 25-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 12176
FILING DATE: 03-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/064,174
FILING DATE: 25-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 016100-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703/836-6620
TELEFAX: 703/836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 579 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-449-733-1

Query Match 4.8%, Score 77.5; DB 2; Length 579;
Best Local Similarity 26.6%; Pred. No. 8.1;
Matches 45; Conservative 26; Mismatches 69; Indels 29; Gaps 9;

QY 76 KATPAGSEGTMRRLAIAFEKPYFKQLMSFYADERSR-----HTVYPPADOVYSTEMCD 130
D 80 RGEPSFSE--RDEL-EKRRGSELLIESKWDGSGRVVGYTNFTYVSGVYLLKNNMD 135
QY 131 IODKVVVLIGDDPYGPNQAHGLCSVQKVPDPPLSLNVIYK---ELCTD-IDGFKHPGH 186
D 136 IKN-NIVLEFGPDGY-----LYIKGKEPSKELPSEKITYTKGTVDYTDAMEKORFESL 186
QY 187 GDLGSMAGQVL--LLNAVLTVRAHQANSHKDRGNET-----FTDAVIK 228
D 187 GSAAGGDKSGALSALEEGVLRLNQAAVASGHTDFGWTSEFEVDFSDKTIK 235

RESULT 14
US-08-487-890A-97
Sequence 97, Application US/08487890A
Patent No. 5708149
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murlin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESSES:

```

```

; ADDRESSEE: Slim & McBurney
; STREET: 6th Floor, 330 University Avenue
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,890A
; FILING DATE: 07-JUN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 97:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 598 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-487-890A-97

```

```

Query Match 4.8%; Score 77.5; DB 1; Length 598;
Best Local Similarity 26.6%; Pred. No. 8.5;
Matches 45; Conservative 26; Mismatches 69; Indels 29; Gaps 9;

```

```

QY 76 KATPAGGEGTWRRLAEEFEKPYFKQLMSFYADERSR-----HTVPPADOVYSTKMD 130
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 100 RGEPNSESE--RDEL--EKRRGSELIESKWEQDSRVGTYNFTYVRSRGVYLYLNKNID 155
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 IODKVVYLGDPPYHGNQAHGLCFVSQKPVPPPSLVNITYK---ELCTD-IDGFKHPGH 186
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 156 IKN-NIVLFEPDGY-----LYYKGEPSKELPSEKITKGTWDVYTDAMEKOREGL 206
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 GDLGMAKQGV--LNLAVLTVRAHQANSKDRGMET----FTDAVIK 228
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 207 GSAAGGKSGALSLAEGVLRNQAEASSGHTDFGMTSEFVDFSDKTIK 255
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 15
; US-08-478-435-97
; Sequence 97, Application US/08478435
; Patent No. 592323
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Hartness, Robin
; APPLICANT: Schuyvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Slim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto

```

```

; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,435
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 97:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 598 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-435-97

```

```

Query Match 4.8%; Score 77.5; DB 2; Length 598;
Best Local Similarity 26.6%; Pred. No. 8.5;
Matches 45; Conservative 26; Mismatches 69; Indels 29; Gaps 9;

```

```

QY 76 KATPAGGEGTWRRLAEEFEKPYFKQLMSFYADERSR-----HTVPPADOVYSTKMD 130
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 100 RGEPNSESE--RDEL--EKRRGSELIESKWEQDSRVGTYNFTYVRSRGVYLYLNKNID 155
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 IODKVVYLGDPPYHGNQAHGLCFVSQKPVPPPSLVNITYK---ELCTD-IDGFKHPGH 186
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 156 IKN-NIVLFEPDGY-----LYYKGEPSKELPSEKITKGTWDVYTDAMEKOREGL 206
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 GDLGMAKQGV--LNLAVLTVRAHQANSKDRGMET----FTDAVIK 228
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 207 GSAAGGKSGALSLAEGVLRNQAEASSGHTDFGMTSEFVDFSDKTIK 255
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Search completed: August 21, 2002, 10:24:56
Job time: 258 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 21, 2002, 10:22:13 ; Search time 17.5 Seconds

(without alignments)
1652.737 Million cell updates/sec

Title: US-09-758-017a-2

Perfect score: 1603
Sequence: 1 MLEFKGLGCRICISSNRVLPG.....SRANGLKLKSGTEPIWMRAL 301Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR.71:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1032	64.4	304	1 A60472	uracil-DNA glycosylase
2	705.5	44.0	228	2 AH0329	probable uracil-DN
3	702.5	43.8	226	2 C82086	uracil-DNA glycosylase
4	683.5	42.6	229	2 DB5904	uracil-DNA glycosylase
5	681.5	42.5	229	1 DGEUC	uracil-DNA glycosylase
6	679.5	42.4	229	2 AH0830	uracil-DNA glycosylase
7	659.5	41.1	218	2 F91059	uracil-DNA glycosylase
8	648.5	40.5	219	2 H81107	uracil-DNA glycosylase
9	635.5	39.6	247	2 A64043	uracil-DNA glycosylase
10	635.5	39.6	247	2 F75486	uracil-DNA N-glyco
11	634.5	39.6	219	2 B81908	probable uracil-DN
12	598	37.3	237	2 AE2878	uracil-DNA glycosylase
13	598	37.3	241	2 G97654	uracil-DNA glycosylase
14	591	36.9	256	2 H82525	uracil-DNA glycosylase
15	586.5	36.6	231	2 C83551	uracil-DNA glycosylase
16	584	36.4	217	2 I36145	uracil-DNA glycosylase
17	572.5	35.7	224	2 B84131	uracil-DNA glycosylase
18	570.5	35.6	225	2 S39712	uracil-DNA glycosylase
19	569	35.5	231	2 F81424	uracil-DNA glycosylase
20	561	35.0	222	2 T40846	uracil-DNA glycosylase
21	559.5	34.9	224	2 AC1228	uracil-DNA glycosylase
22	559.5	34.9	224	2 AE1581	uracil-DNA glycosylase
23	557	34.7	218	2 G89826	uracil-DNA glycosylase
24	552	34.4	229	2 B81652	uracil-DNA glycosylase
25	548	34.4	220	2 DB4951	uracil-DNA glycosylase
26	541	34.2	229	2 C71493	probable uracil DN
27	546.5	34.1	221	2 AG1123	uracil-DNA glycosylase
28	540	33.7	359	1 A31425	uracil-DNA glycosylase
29	539.5	33.7	236	2 C86587	uracil DNA glycosylase

30	539.5	33.7	236	2 H72035	uracil-DNA glycosylase
31	536.5	33.5	221	2 A11483	uracil-DNA glycosylase
32	535.5	33.4	217	2 E95135	uracil-DNA glycosylase
33	533.5	33.3	217	2 S13591	uracil-DNA glycosylase
34	533.5	33.3	222	2 G98003	DNA-uracil glycosylase
35	496.5	31.0	223	2 E70106	uracil DNA glycosylase
36	474	29.6	314	2 T42604	probable uracil-DN
37	472.5	29.5	312	1 DGBE83	uracil-DNA glycosylase
38	472	29.4	252	2 T03140	uracil-DNA glucosylase
39	463.5	28.9	281	2 S33023	probable uracil-DN
40	457	28.5	255	2 S55641	uracil DNA glycosylase
41	454.5	28.4	219	2 B86654	uracil-DNA glycosylase
42	451	28.1	252	2 T42961	probable uracil-DN
43	446.5	27.9	252	2 G62877	uracil-DNA glycosylase
44	445	27.8	252	1 DGBEN8	uracil-DNA glucosylase
45	429.5	26.8	250	1 DGBEL5	uracil-DNA glycosylase

ALIGNMENTS

```
RESULT 1
A60472
uracil-DNA glycosylase (PC 3.2.2.-) precursor [validated] - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 08-Dec-2000
C:Accession: S05964; A60472
R:Olson, L.C.; Aasland, R.; Wiltner, C.U.; Krokan, H.E.; Helland, D.E.
EMBO J. 8, 3121-3125, 1989
A:Title: Molecular cloning of human uracil-DNA glycosylase, a highly conserved DNA re
A:Reference number: S05964; MUID:90059899
A:Accession: S05964
A:Molecule type: mRNA
A:Residues: 1-304 <OLS>
A:Cross-references: EMBL:X15653; NID:g37598; PIDN:CAA33679.1; PID:g37599
R:Wiltner, C.U.; Bauw, G.; Krokan, H.E.
Biochemistry 28, 780-784, 1989
A:Title: Purification and determination of the NH-2-terminal amino acid sequence of u
A:Reference number: A60472; MUID:89229080
A:Accession: A60472
A:Molecule type: Protein
A:Residues: 78-93, 'X', 95-104 <MIT>
C:Genetics:
A:Gene: GDB:UNG, DGU, UDG
A:Cross-references: GDB:119844; OMIM:191525
A:Map position: 12pter-12qter
C:Superfamily: uracil-DNA glycosylase
C:Keywords: DNA repair; glycosylase; hydrolase; mitochondrion
F:1-77/Domains: transit peptide (mitochondrion) #status predicted <TNP>
F:78-304/Product: uracil-DNA glycosylase #status experimental <MAT>

Query Match 64.4%; Score 1032; DB 1; Length 304;
Best Local Similarity 67.4%; Pred. No. 6-3e-81;
Matches 194; Conservative 32; Mismatches 56; Indels 6; Gaps 3;

QY 19 PG---LLPPTLCFSKLMKTPKLRSSNVEQKT--SSQLSVEDELERNAKKKKAAALDKI 73
      || | : || | | | | | | | | | | | | | | | | | | | | | | | | |
DB 18 PKGPIQLLSRLCGDHLQAIIPAKKAPAGOEPEPTPSSP-LSAEQIDRIORAKKAALLRL 76
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 74 RAKATPAGGEGTWRKRELAEEFEPYFKOLMSFVADERSNHTYTPPADQVYSSTEMCDDIOD 133
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 77 AARNPVVPGESGEMKRLDSEGFEPYFIKMGFVAERKHYTVYPPHOVFTQMCDDID 136
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 VVVVILGDPYHGPQNAHGLCFSVOKPVPVPSLVNIYKELCTDIDGFKHPHGLDSGMA 193
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 137 VVVVILGDPYHGPQNAHGLCFSVQRPVPPPSLENIYKELSTDIEDYFHPHGLDSGMA 196
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 194 KGGVLLNNAVLTVRHAQANSKDRGMEFTTDAVIRKLSVNRGGVYVFLMGSTAAHKKGATV 253
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 197 KGGVLLNNAVLTVRHAQANSKDRGMEFTTDAVIRKLSVNRGGVYVFLMGSTAAHKKGATV 256
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 254 DKRRHNVLDQAVHPSPLSAHRGFLGCKHFKSKANGLKLKSTDEPIWMRAL 301
```

Db 257 DRKRHHVLOTAAHPSPLSVYRGFGCRHFSKTNELLQKSGKKPIDMKEL 304

RESULT 2

AH0329

Probable uracil-DNA glycosylase (EC 3.2.2.-) [Imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001

C:Accession: AH0329

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarrera, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AH0329

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-228 <KUR>

C:Cross-references: GB:AL590842; PIDN:CAC92943.1; PID:g15980682; GSPDB:GN00175

C:Genetics:

A:Gene: ung

C:Superfamily: uracil-DNA glycosylase

C:Keywords: glycosidase; hydrolase

Query Match 44.0%; Score 705.5; DB 2; Length 228;
Best Local Similarity 60.2%; Pred. No. 4.6e-53;
Matches 130; Conservative 34; Mismatches 51; Indels 1; Gaps 1;

85 TWRRLEAEFEKPYFKOLMSFVADER-SRHTVVPADQVYSTKCDIQDKVVLGGDP 143

Db 6 TWHDVIGCEKQPYFKDLAVAAERAKTIYPPQKIDFNAFRLTELDQYKVVILGGDP 65

QY 144 YHGRNQAAGLCFSYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMAGQVLLNAV 203

Db 66 YHGRNQAAGLCFSYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMAGQVLLNAV 125

QY 204 LTVRAHQANSKDKGMEFTDAVTKLSVNRGVFLLMGSAVANKKATIDRKHHVLOA 263

Db 126 LTVRGAQSHASLNGWEFTTKVIALNENHREGVIFMLGSHAKKGRILNTERHHVILKA 185

QY 264 VHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 299

Db 186 PHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 221

RESULT 3

C82086

uracil-DNA glycosylase VC2359 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: C82086

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, H.

1, R.R.; Mkalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: C82086

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-226 <HEI>

A:Cross-references: GB:AE004306; GB:AE003852; NID:9656924; PIDN:AAF95502.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2359

A:Map position: 1

C:Superfamily: uracil-DNA glycosylase

Query Match 43.8%; Score 702.5; DB 2; Length 226;

Best Local Similarity 61.1%; Pred. No. 8.2e-53;
Matches 132; Conservative 25; Mismatches 58; Indels 1; Gaps 1;

85 TWRRLEAEFEKPYFKOLMSFVADER-SRHTVVPADQVYSTKCDIQDKVVLGGDP 143

Db 6 TWHDVIGCEKQPYFKDLAVAAERAKTIYPPQKIDFNAFRLTELDQYKVVILGGDP 65

QY 144 YHGRNQAAGLCFSYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMAGQVLLNAV 203

Db 66 YHGRNQAAGLCFSYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMAGQVLLNAV 125

QY 204 LTVRAHQANSKDKGMEFTDAVTKLSVNRGVFLLMGSAVANKKATIDRKHHVLOA 263

Db 126 LTVRGAQSHASLNGWEFTTKVIALNENHREGVIFMLGSHAKKGRILNTERHHVILKA 185

QY 264 VHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 299

Db 186 PHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 221

RESULT 4

D85904

uracil-DNA-glycosylase [Imported] - Escherichia coli (strain O157:H7, substrain EDL93

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: D85904

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

11ler, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: D85904

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <STO>

A:Cross-references: GB:AE005174; NID:912517001; PIDN:AAG57696.1; GSPDB:GN00145; UMG:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ung

C:Superfamily: uracil-DNA glycosylase

Query Match 42.6%; Score 683.5; DB 2; Length 229;
Best Local Similarity 60.9%; Pred. No. 3.6e-51;
Matches 131; Conservative 24; Mismatches 59; Indels 1; Gaps 1;

85 TWRRLEAEFEKPYFKOLMSFVADER-SRHTVVPADQVYSTKCDIQDKVVLGGDP 143

Db 6 TWHDVIGCEKQPYFKDLAVAAERAKTIYPPQKIDFNAFRLTELDQYKVVILGGDP 65

QY 144 YHGRNQAAGLCFSYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMAGQVLLNAV 203

Db 66 YHGRNQAAGLCFSYQKRPVPPPSLVNLYKELCTDIDGFKHGHGDLGSMAGQVLLNAV 125

QY 204 LTVRAHQANSKDKGMEFTDAVTKLSVNRGVFLLMGSAVANKKATIDRKHHVLOA 263

Db 126 LTVRGAQSHASLNGWEFTTKVIALNENHREGVIFMLGSHAKKGRILNTERHHVILKA 185

QY 264 VHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 298

Db 186 PHPSPLSAHRGFLGCKHPSKANGLIKLSGTEPIMR 220

RESULT 5

D8ECU

uracil-DNA glycosylase (EC 3.2.2.-) - Escherichia coli

C:Species: Escherichia coli

C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000

C:Accession: A28175; C65036

R:Varshney, U.; Hutcheon, T.; van de Sande, J.H.

J. Biol. Chem. 263, 7776-7784, 1988

A:Title: Sequence analysis, expression, and conservation of Escherichia coli uracil D

A:Reference number: A28175; MUID:88227981

A:Accession: A8175
A:Molecule type: DNA
A:Residues: 1-229 <VAR>
A:Cross-references: GB:D64044; NID:g987635; PIDN:BA10923.1; PID:g987651
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: C65036
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-229 <BLAT>
A:Cross-references: GB:A000344; GB:U00096; NID:q1788927; PIDN:AAC75633.1; PID:q1788934
A:Experimental source: strain K-12, substrain MG1655
C:Comment: This enzyme catalyzes the hydrolysis of the deoxyribose N-glycosidic bond to
on.
C:Genetics:
A:Gene: ung
A:Map position: 56 min
C:Superfamily: uracil-DNA glycosylase
C:Keywords: DNA repair; glycosidase; hydrolase

```
Query Match      42.5% Score 681.5; DB 1; Length 229;  
Best Local Similarity    60.9%; Pred. No. 5.3e-51;  
Matches 131; Conservative   24; Mismatches 59; Indels 1; Gaps 1;
```

Oy 85 TWRRRLAEFFKPFYKOALMSFVADER-SRHTVTYPADVOYSTEMCDIQDYKVITLIGDP 143
 | :
Db 6 TWHVDLAEEKQPTLYNLDTVASEROSGVITIIPPKKYVFNAFRTELDGDKAVYLIGDP 65

Oy 144 YHGNQAHGILCSYOKPVPPPSLVNIYKELTCTDIDGFKNHBGDLSGMAROGVLLNAV 203
 ||||| :
Db 66 YHGPOAGHLAFSVPRGAIRPSSLINMYKELENTIFGTNRNHGYLESMAROGVLLNTV 125

Oy 204 LTVRAHQNSAKHDGEWTFDAVKMLSLVNREGVVFLMSGVAHKKGATIDRKRNHYIDA 263
 ||||| :
Db 126 LTVRGGQHSHASLGWEFTETDKVISLINOHREGVVELLMGSAAKKKAIDIKRHIVHLA 185

Oy 264 VHPSPSAHRGFLCGKHFSKANGLKTSGTPIPW 298
 ||||| :
Db 186 PHSPLSABRGFGFCNHFVLANOMLEORNGERTPIDW 220

RESULT 6
uracil-DNA glycosylase (EC 3.2.2.-) [imported] - Salmonella enterica subsp. enterica serovar Typhimurium strain DT104
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: This species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH0830
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Parkhill, J.; Croxall, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
J.S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001.

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AH0830
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1..229 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05831.1; PID:gI6503806; GSPDB:GN00176
C:Genetics:
A:Gene: STR2840
C:Superfamily: uracil-DNA glycosylase
C:Keywords: glycosidase; hydrolase

Query Match 42.4% Score 679.5; DB 2; Length 229;
Best Local Similarity 60.9%; Pred. No. 7.9e-51;
Matches 131; Conservative 24; Mismatches 59; Indels 1; Gaps 1;

[illegible]

RESULT

7

uracil-DNA-glycosylase ECS3446 [imported] - Escherichia coli (strain O157:H7, substrate P91059)

C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: F91059

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kuahara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res.: 8, 11-22, 2001

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F91059
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1218 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAB36869.1; PID:g13362917; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RTMD 0509952
C:Genetics:
A:Gene: ECS3446
C:Superfamily: uracil-DNA glycosylase

Query Match **41.1%**; Score 659.5; DB 2; Length 218;

Best Local Similarity 61.7%; Pred. No. 3.8e+49;

Matches 127; Conservative 23; Mismatches 55; Indels 1; Gaps 1;

OY 85 TWRRRLAEFEKPYRKQIMSFVADER-SRHYVPYPADVOYSTECMIDQDKVVYLGGDP 143
| | | : : : : : | | | :
Db 6 TWDVLAEKEOKOPYFLNTLQTIVASEROGSVITIPPQKKVFNAFRFTGADVKNVILGGP 65
OY 144 YHGPNQAGLGLEFSYQKPAPPPESLAVNIYEKLCTDIDGFGRHGSDLSGMAKOGVLLNAV 203
| | | | | : : : : : | | | : : : : : | | | | | : : : : : : : : : : : : : :
Db 66 YHGPGQAGLAFVSVRPATRPSSILLNMKYCLEINTIPGTTRRNHGGLYESMAQGVLLNTV 125
OY 204 LTVRAHQANSKRDCWETFDVAIKWLISVNREGVFFLIWSYAHKKATIDRKRHVQLQA 263
| | | | | : : : : : | | | : : : : : | | | | | : : : : : : : : : : : : : :
Db 126 LTVRAGQASHASLASLMETFPTKDVISLINQHREGVFLLMSGHAQKCALIDKORHVLA 185
OY 264 VHPSPLSAHRGPFGCKHKRSKANGLK 289
| | | | | : : : : : | | | : : : : : | | | | | : : : : : : : : : : : : : :
Db 186 PHPSPLSAHRGPFGCNHFVLANOWLE 211

RESULT

8

uracil-DNA glycosylase NMBI222 [imported] - Neisseria meningitidis (strain MC58 serog
P91107)

C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: HB1107

R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
Li, H.; Qin, H.; Yamdathan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappapoli, R.;
Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

```

A:Reference number: A81000; MUID:20175755
A:Accession: H81107
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-219 <TEF>
A:Cross-references: GB:AEO02470; GB:AEO02098; NID:g7226457; PIDN:AAFA1604.1; PID:g7226444
A:Experimental source: serogroup B, strain KC58
C:Genetics:
C:Gene: NM81222
C:Superfamily: uracil-DNA glycosylase

Query Match      40.5%; Score 648.5; DB 2; Length 219;
Best Local Similarity 55.8%; Pred. No.3,4e-47;
Matches 121; Conservative 35; Mismatches 60; Indels 1; Gaps 1;

OY 84 ETWRRELAEEFKRPFYKOLMSFVADER-SRHITYVPRAQOVYSTEMCDIQDKVYLIGOD 142
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||
DB 2 DTHWHAALGGEKKQPPRFQEILNAVROERLSGOITTPRADVFARFLTLAFDRKAVIIGOD 61
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||

OY 143 PYHGPNQAHGCLCFYQVKRPVPPPSLVNTIKELCTDIDEGFKHGHGDLDSGMAKOGVLLNTA 202
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||
DB 62 PYHGAGOAHGALFASRYKRGIRIPPSLINTYKELETDEGFSIPAHGCLTAMAEQGVLLNT 121
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||

OY 203 VLTVAHQANSHKDGMEFTFDAYIKMLSVNREGVFFLLMSGYAANKGATIDRKRNHHVLQ 262
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||
DB 122 VLTVRGAQHSHALLGWERTDTVIROLATHRKLHLVFMLMGVDAQKGRLIDSQNHLILT 181
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||

OY 263 AVHPSPLSAHRGFLGCKHFESKANGLLKSTGTEPIIMNR 299
   ||||||:::|||||:::|||||:::|||||
DB 182 APHPSPLSAHRGFSGCRHSQSANSYLSKRHDIPIMWK 218
   |||||||:::|||||:::|||||:::|||||

RESULT 9
A64043
uracil-DNA glycosylase (BC 3.2.2.-) - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 22-Jun-1999
C:Accession: A64043
R:Flitschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gaynes, J.D.; Scott, J.; Shibley, R.; Liu, L.I.; Glodet, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fritschman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A. Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. A:Reference number: A64000; MUID:95350630
A:Accession: A64043
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-219 <TIGR>
A:Cross-references: GB:U32687; GB:I42023; NID:g15729955; PIDN:AAC21696.1; PID:g1572962; C:Superfamily: uracil-DNA glycosylase
C:Keywords: DNA repair; glycosidase; hydrolase

Query Match      39.6%; Score 635.5; DB 2; Length 219;
Best Local Similarity 57.1%; Pred. No.4,4e-47;
Matches 124; Conservative 28; Mismatches 64; Indels 1; Caps 1;

OY 84 ETWRRELAEEFKRPFYKOLMSFVADER-SRHITYVPRAQOVYSTEMCDIQDKVYLIGOD 142
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||
DB 2 KNWDIVIGTEKAQPFQHTLOVNLARASGKITYPPOEDVNARKYTAFEDVKVILLIGOD 61
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||

OY 143 PYHGPNQAHGCLCFYQVKRPVPPPSLVNTIKELCTDIDEGFKHGHGDLDSGMAKOGVLLNTA 202
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||
DB 62 PYHGNOAHGALFASRYKREVALPPLSLNTYKELETDISGFQMPNSGYLTKKMAEQGVLLNT 121
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||

OY 203 VLTVAHQANSHKDGMEFTFDAYIKMLSVNREGVFFLLMSGYAANKGATIDRKRNHHVLQ 262
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||
DB 122 VLTVRGAQHSHANLGWERFTDKVIADVLENHREKLFYFLMWSHNOKKOMIDRRRHVLT 181
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

[illegible]

Query Match	36.9%;	Score 591;	DB 2;	Length 256;
Best Local Similarity	53.2%;	Pred. No. 3.6e-43;		
Matches 116;	Conservative 31;	Mismatches 63;	Indels 8;	Gaps 3;

QY	85	TWRRLAEFEKPYKOLMSFVADERSRH----	TYPPADQVYSTECMDIODYVVLIG	140
		: : : : : : : : : :		
Db	24	SWKAVHGVWMLRPEKRDLSAFL---	RARKVAGVSYTPPGSOLFALFEATPFOKKAIVLIG	80
QY	141	QDPYHGPNQAHGLCSFVQKPYPPPSILVNIYKELCTDIDGCKHPHGSLDSMAKOGVLLL		200
		: : : : : : : : : :		
Db	81	QDPYHGQOAHGLCSFVAPGMPLPSLINTYKELDEEDL-GLLRPDHGCLLPMARKGVLLL		139
QY	201	NAVLTVYRAHQANSKDRQWETFTDAVITKMLSVNREGGVFLLMGSTAHKKKCATIDRKRHHV		260
		: : : : : : : : : :		
Db	140	NAVLTVEDEGRGAHOGKMGEGFTDHYVDTLNNEBGLVFMJMGSSYAQAKGVITDRRLHV		199
QY	261	LOAVHPSPLSARHGLGCKHFSKANGLLKSGTEPIINM		298
		: : : : : : : : : :		
Db	200	LKAAPSPLSAHRGFLGCRHFSLCLNQVYLSQHGLGAWDM		237

RESULT 15

uracil-DNA glycosylase PA0750 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: C83551
 R:Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Br
 adman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., L
 iory, S., Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
 A:Reference number: AB2950; MUID:20437337
 A:Accession: C83551
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-231 <STO>
 A:Cross-references: GB:AE004510; GB:AE004091; NID:g9946634; PIDN:AA00139.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: ung; PA0750
 C:Superfamily: uracil-DNA glycosylase

Query Match	36.6%;	Score 586.5;	DB 2;	Length 231;
Best Local Similarity	49.8%;	Pred. No. 7.6e-43;		
Matches 114;	Conservative 40;	Mismatches 66;	Indels 9;	Gaps 3;

QY	71	DKRAKATPAPGEMWREBELAAEFKEPKYKOLMSFVADBSER-HTVYPRADVOYSSTRMC	12
Db	6	DIKLEA-----SWKALBEEEDKPKMKOLGEELEKROEKAGKAVITPEPGLFIPNALNTT	58
QY	130	DIDQKVVILGODPYHCGNQAHGICFSYQKCVPPPEPSLVNAYKELCDIDGFKHPRGDL	189
Db	59	PLENKVVITIDOPPHGCGAHGICFSYQPEVPPIPSLQNTIKRLNDLN-IPINNIGYL	117
QY	190	SGMAKGVLLINAVLTJVRAHQANSHKDRGWETFFDAVITKWLSVNREGVETLLMGSYAHK	249
Db	118	QMAEOGVLLINTSLTVEQAKAGSHANAGMOPETFDRIYEVVNERCERLIVELMGSSHAK	177
QY	250	GATDRKRRHHVLOAVHPSPLSAHSGICGCKHFSKANGILKLSGTFEPINW	298
Db	178	QKLDPOHHLIKSAHPSPLATYKGFUNGHSSTNFKLEONGKGTPIDW	226

Search completed: August 21, 2002, 10:25:17
Job time: 184 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 21, 2002, 10:24:58 ; Search time 13.44 Seconds

(without alignments)
867.157 Million cell updates/sec

Title: US-09-758-017a-2

Perfect score: 1603

Sequence: 1 MFKLGICORCISNRVLPG.....SKANGLLKLGTEPINRAL 301

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1051.5	65.6	295	1 UNG_MOUSE	P97931 mus musculu
2	1032	64.4	304	1 UNG_HUMAN	P13051 homo sapien
3	681.5	42.5	228	1 UNG_ECOLI	P12295 escherichia
4	665.5	41.5	222	1 UNG_PASMU	P57807 pasteurella
5	635.5	39.6	219	1 UNG_HAETN	P33731 haemophilus
6	601	37.5	257	1 UNG_DICDI	P33766 dictyosteli
7	597	37.2	217	1 UNG_PSEDE	P29950 pseudomonas
8	572.5	35.7	224	1 UNG_BACHD	O9K682 bacillus ha
9	570.5	35.6	225	1 UNG_BACSU	P19615 bacillus su
10	561	35.0	322	1 UNG_SCHRO	O74834 schizosach
11	552	34.4	229	1 UNG_CHLMU	O9P342 chlamydia m
12	551	34.4	220	1 UNG_BUCAI	P57280 buchera ap
13	548	34.2	229	1 UNG_CHLFR	O84613 chlamydia t
14	540	33.7	236	1 UNG_YEAST	P12887 saccharomyc
15	539.5	33.7	236	1 UNG_CHLPN	O92743 chlamydia p
16	537.5	33.5	217	1 UNG_STRAG	O9X468 streptococc
17	535.5	33.4	217	1 UNG_STRPN	P23379 streptococc
18	496.5	31.0	223	1 UNG_BORBU	O51082 borrelia bu
19	472.5	29.9	312	1 UNG_HSVB	P18866 equine herp
20	463.5	28.9	255	1 UNG_EBV	P12888 Epstein-Bar
21	457	28.5	255	1 UNG_HSV2	P53755 equine herp
22	454.5	28.4	219	1 UNG_LACIA	O9C122 lactococcus
23	454.5	28.4	315	1 UNG_PRTVE	P52506 pseudorabie
24	454.5	28.4	316	1 UNG_PPRVA	O01019 pseudorabie
25	445	27.8	252	1 UNG_HSVSA	O01019 herpesvirus
26	429.5	26.8	250	1 UNG_HCVVA	P16769 human cytom
27	423	26.4	334	1 UNG_HSV1	P10166 herpes simp
28	422	26.3	255	1 UNG_HSV2H	P58275 herpes simp
29	415	25.9	255	1 UNG_HSV6U	P52345 human herpe
30	411.5	25.7	204	1 UNG_HSVBC	P53764 bovine herp
31	411	25.6	255	1 UNG_HSV62	P52447 human herpe
32	408.5	25.5	254	1 UNG_HSV7J	P50639 human herpe
33	401	25.0	294	1 UNG_HSV23	P13158 herpes simp

ALIGNMENTS

RESULT	1	STANDARD	PRT	295 AA.	
UNG_MOUSE					
AC	P97931; P97509;				O92j99 helicobacte
DT	01-NOV-1997 (Rel. 35, Created)				P09307 varlicella-z
DT	01-NOV-1997 (Rel. 35, Last sequence update)				P56397 helicobacte
DT	16-OCT-2001 (Rel. 40, Last annotation update)				P95119 mycobacteri
DE	Uracil-DNA glycosylase, mitochondrial precursor (EC 3.2.2.-) (UDG).				O9Cbs3 mycobacteri
GN	UNG OR UNG1.				P75536 mycoplasma
OS	Mus musculus (Mouse).				P47343 mycoplasma
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				P21968 fowlpox vlr
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				O10231 schizosach
OX	NCBI_TaxID=10090;				O99490 homo sapien
RN	[1]				O61909 mus musculu
RP	SEQUENCE FROM N.A.				P04303 vaccinia vi
RX	MEDLINE=97311407; PubMed=9168124;				
RA	Synden P.C., Yee H.A., Winkeln R.J., van de Sande J.H.;				
RT	"The mouse uracil-DNA glycosylase gene: isolation of cDNA and genomic clones and mapping ung to mouse chromosome 5.";				
RL	Gene 189:175-181(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97169285; PubMed=9016624;				
RA	Nilsen H., Solum K., Haug T., Krokan H.E.;				
RT	"Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the ung gene.";				
RL	Nucleic Acids Res. 25:750-755(1997).				
CC	-1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE.				
CC	-1- SUBCELLULAR LOCATION: NUCLEAR AND MITOCHONDRIAL (BY SIMILARITY).				
CC	-1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U55040; AAB39511.1; -;				
DR	EMBL; U55041; AAC53197.1; -;				
DR	EMBL; X99018; CAA67489.1; -;				
DR	HSSP; P13051; IAKZ.				
DR	MCD; MG1:109352; Ung.				
DR	InterPro: IPR003243; U_DNA_glycosylase.				
DR	InterPro: IPR003249; Urac_DNA_glycosyl.				
DR	Pfam; PF00315; UNG; 1.				
DR	PRODOM; PD001589; Urac_DNA_glycosyl; 1.				
DR	PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.				
KW	DNA repair; Hydrolase; Glycosidase; Nuclear protein; Mitochondrion; Transil peptide.				
FT	TRANSIT 1 58 MITOCHONDRION (POTENTIAL).				

FT	CHAIN	59	295	URACIL-DNA GLYCOSYLASE.
FT	ACT SITE	136	136	GENERAL BASE (BY SIMILARITY).
FT	CONFLICT	266	266	H -> Y (IN REF. 2).
SO	SEQUENCE	295 AA;	33054 MW;	7665656DEC35B851 CRC64;
Query Match				
Best Local Similarity		65.6%;	Score 1051.5;	DB 1; Length 295;
Matches 191;		Conservative	32;	Mismatches 47; Indels 9; Gaps
QY	32 LMKITP-----	KKLRSSNVDEKTS	SSPOLSVDELRMAKNKAALDKIRAKATPAG	82
Db	17 LRSLEPNDDSDRSRQAPAKKARVEONEGSPLEAQLVRIORNKAAALLRLAARNVAG	76		
QY	83 GETWRRELAAEFERKPYFQMLSFVADESRRTVVPADQVYSSTEMCDIDQVKVILGOD	14		
Db	77 GESWKOQJCGEGRGKRYFKLMGFVAEEENHNRKVPYPPPEVFTWQMDIRVKVILGOD	130		
QY	143 PYHGPNQAHGICFSYQKRVPPPSLVNTYKELCDIDDFKHPHGHDLSGMAKOGVLLNA	20		
Db	137 PYHGPNQAHGICFSYQKRVPPPSLVNTYKELCDIDDFKHPHGHDLSGMAKOGVLLNA	199		
QY	203 VLTVARAHQANSHKDRGEMETFFDAVYIKWLVSYBREGVFTLLSGSYAHKKGATIDRRHHVLY	26		
Db	197 VLTVARAHQANSHKDRGEMETFFDAVYIKWLVSYBREGVFTLLSGSYAHKKGATIDRRHHVLY	256		
QY	263 AVHPSPLSAHGFLGCKHFKSANGKLLKLSGNETPWRML	301		
Db	257 TAHPSPLSVHKGFLGCKHFKSANGKLLKLSGNETPWRML	295		
RESULT 2				
UNG_HUMAN	UNG_HUMAN	STANDARD;	PRT;	304 AA.
AC	P13051:			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Uracil-DNA glycosylase, mitochondrial precursor (EC 3.2.2.-) (UDG).			
GN	OR DGU OR UNG15.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RN	SEQUENCE FROM N.A. AND SEQUENCE OF 78-104.			
RC	TISSUE=Placenta;			
RP	MEDLINE=90059899; Pubmed=255154;			
RA	Olsen L.C., Aasland R., Wittwer C.U., Krokan H.E., Helland D.E.;			
RT	"Molecular cloning of human uracil-DNA glycosylase, a highly			
RL	conserved DNA repair enzyme."			
RL	EMBO J. 8:3121-3125(1989).			
RN	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=95010790; Pubmed=7926048;			
RA	Haug T., Skorpjen F., Lund H., Krokan H.E.;			
RT	"Structure of the gene for human uracil-DNA glycosylase and analysis			
RL	of the promoter function."			
RL	FEBS Lett. 353:180-184(1994).			
RN	[3]			
RP	SUBCELLULAR LOCATION.			
RP	MEDLINE=93324318; Pubmed=8332455;			
RA	Slupphaug G., Markussen F.-H., Olsen L.C., Aasland R., Aarsaether N.,			
RA	Bakke O., Krokan H.E., Helland D.E.;			
RT	"Nuclear and mitochondrial forms of human uracil-DNA glycosylase are			
RT	encoded by the same gene."			
RL	Nucleic Acids Res. 21:2579-2584(1993).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).			
RP	MEDLINE=95211838; Pubmed=7697717.			
RA	Mol C.D., Arvai A.S., Slupphaug G., Kavli B., Alseth I., Krokan H.E.,			
RA	Tainer J.A.;			
RT	"Crystal structure and mutational analysis of human uracil-DNA			

RT	glycosylase: structural basis for specificity and catalysis."
RN	Cell 80:869-878(1995).
RN	[5]
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX	MEDLINE=95401260; PubMed=7671300;
RA	Mol C.D., Arvai A.S., Sanderson R.J., Slupphaug G., Kavli B.,
RA	Krokan H.E., Moshagen D.W., Tainer J.A.;
RT	"Crystal structure of human uracil-DNA glycosylase in complex with a
RT	protein inhibitor: Protein mimicry of DNA.";
RL	Cell 82:701-708(1995).
RN	[6]
RN	X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RP	MEDLINE=97055940; PubMed=8900285;
RX	Slupphaug G., Mol C.D., Kavli B., Arvai A.S., Krokan H.E.,
RA	Tainer J.A.;
RA	"A nucleotide-flipping mechanism from the structure of human
RT	uracil-DNA glycosylase bound to DNA.";
RL	Nature 384:87-92(1996).
RN	[7]
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 85-304.
RX	MEDLINE=98393562; PubMed=9724657;
RA	Patikh S.S., Mol C.D., Slupphaug G., Bharrati S., Krokan H.E.,
RA	Tainer J.A.;
RT	"Base excision repair initiation revealed by crystal structures and
RT	binding kinetics of human uracil-DNA glycosylase with DNA.";
RL	EMBO J. 17:5214-5226(1998).
RN	[8]
RP	MUTAGENESIS.
RX	MEDLINE=96283644; PubMed=8670846;
RA	Kavli B., Slupphaug G., Mol C.D., Arvai A.S., Petersen S.B.,
RA	Tainer J.A., Krokan H.E.;
RT	"Excision of cytosine and thymine from DNA by mutants of human
RT	uracil-DNA glycosylase.";
RL	EMBO J. 15:3442-3447(1996).
CC	-1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC	AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC	POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE.
CC	-1- SUBUNIT: MONOMER.
CC	-1- SUBCELLULAR LOCATION: Nuclear and mitochondrial.
CC	-1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X15653; CAA33679.1; -
DR	EMBL; X89398; CAA61579.1; -
DR	PIR; S05964; S05964.
DR	PDB; 1AKZ; 20-AUG-97.
DR	PDB; 1UGH; 09-FEB-99.
DR	PDB; 4SKN; 26-FEB-99.
DR	PDB; 1SSP; 06-MAY-99.
DR	PDB; 2SSP; 06-MAY-99.
DR	MIM; 191525; -
DR	InterPro; IPR002043; U_DNA_glycosylase.
DR	InterPro; IPR003249; Urac_DNA_glycosyl.
DR	Pfam; PF00315; UNG; 1.
DR	ProDom; PD001589; Urac_DNA_glycosyl; 1.
DR	ProSite; PS00130; U_DNA_GLYCOSYLASE; 1.
KW	DNA repair; Hydrolase; Glycosidase; Nuclear protein; Mitochondrion;
FT	Transit peptide; 3d-structure.
KW	Transit 1
FT	77
FT	CHAIN 78
FT	ACT SITE 145
FT	MUTAGEN 145
FT	MUTAGEN 147
FT	204
SEQUENCE	304 AA; 33924 MW; 32998C244EA7B215 CRC64;

```

Query Match          64.4%; Score 1032; DB 1; Length 304;
Best Local Similarity 67.4%; Pred. No. 2,5e-84;
Matches 194; Conservative 32; Mismatches 56; Indels 6; Gaps

OY 19 PG-----LLPQLCSKLMKIKPKILRSSVNGK*-SSPOLSVQLERMKKNKAALDKI 73
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 18 PGRKPLQLSLKLCGDHLDQATPAKKAAPQGEPRGPSP-LSAQDLRIORKNKAALLRL 76
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 74 RAKATPAGCGETWRRELEAAFEKRPYKQLMSFVADERSRHTYVPPADQVSYSTEMCIDI 133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 77 AARNVPVGGESGMKKHLISGEGKRPYFIKLMGFVAEEKRHTYVPPHQVFTWQMCDIKD 136
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 134 VKVYILGDDPRVHGNGQAHGLCFESYOKRPVPPRLVNYIKELCTIDGFKRPHGDLSGMA 133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 137 VKVYILGDDPRVHGNGQAHGLCFESYOKRPVPPRLVNYIKELCTIDGFKRPHGDLSGMA 166
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 194 KQGVLLNAVLTVRAHQANSRHKDGWETFDYAVIKWLVSNGEVFLFMGSYAHKKGATI 253
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 197 KQGVLLNAVLTVRAHQANSRHKDGWETFDYAVIKWLVSNGEVFLFMGSYAHKKGATI 256
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 254 DRKRHHVLQAVHPSPSLAHKRGFLCCKHFPSKANGLLKLSGTEPIWRKL 301
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 257 DRKRHHVLQAVHPSPSLAHKRGFLCCKHFPSKANGLLKLSGTEPIWRKL 304
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
UNG_ECOLI STANDARD; PRT; 228 AA.
AC P12295;
AD 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Uricil-DNA glycosylase (EC 3.2.2.-) (UDG).
GN UNG OR B2580.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RP SEQUENCE FROM N.A., AND SPOUNCE OF 1-31.
RX MEDLINE=68227981; Pubmed=2836397;
RA Varshney U., Hutcheon T., de Sande J.H.;
RT "Sequence analysis, expression, and conservation of Escherichia coli
RT uracil-DNA glycosylase and its gene (ung).";
RL J. Biol. Chem. 263:7776-7784(1988).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-K12;
RC "Non-ribosomal proteins affecting the assembly of ribosomes in
RA Escherichia coli.";
RT (In) Nierhaus K.H. (eds.);
RL The translational apparatus, pp.185-195, Plenum Press,
RN New York (1993).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN-K12;
RC Nashimoto H., Saito N.;
RA Nashimoto H., Saito N.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RX MEDLINE=97426617; Pubmed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[5]
RN SEQUENCE FROM N.A.
RP

```

RC STRAIN-K12: PubMed-9205837.
RX MEDLINE=97349980. PubMed-9205837.
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampel G., Satoh Y., Sivasubraman S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.,
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [6]
RP STRUCTURE BY NMR.
RX MEDLINE=97407932. PubMed-9261156.
RA Lundquist A.J., Beger R.D., Bennett S.E., Bolton P.H., Mosbaugh D.W.,
RT "Site-directed mutagenesis and characterization of uracil-DNA
RT glycosylase inhibitor protein. Role of specific carboxylic amino
RT acids in complex formation with Escherichia coli uracil-DNA
RT glycosylase.";
RL J Biol. Chem. 272:21408-21419(1997).
RN [7]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=99451580. PubMed-9776748;
RA Ravishanker R., Sagar M.B., Roy S., Purnapatre K., Handa P.,
RA Varshney U., Vijayan M.,
RT "X-ray analysis of a complex of Escherichia coli uracil DNA
RT glycosylase (EcUDG) with a proteinaceous inhibitor. The structure
RT elucidation of a prokaryotic UDG.";
RN Nucleic Acids Res. 26:4880-4887(1998).
RL [8]
RP X-RAY CRYSTALLOGRAPHY (1.55 ANGSTROMS).
RX MEDLINE=99182421. PubMed=10080896;
RA Putnam C.D., Thayer M.J.N., Lundquist A.J., Mol C.D., Arvai A.S.,
RA Mosbaugh D.W., Trayner J.A.,
RT "Protein mimicry of DNA from crystal structures of the uracil-DNA
RT glycosylase inhibitor protein and its complex with Escherichia coli
RT uracil-DNA glycosylase.";
RL J. Mol. Biol. 287:331-346(1999).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX MEDLINE=99188668. PubMed=10090282;
RA Xiao G., Tordova M., Jagadeesh J., Drohat A.C., Stivers J.T.,
RA Gilliland G.L.,
RT "Crystal structure of Escherichia coli uracil DNA glycosylase and its
RT complexes with uracil and glycerol: structure and glycosylase
RT mechanism revisited.";
RL proteins 35:13-24(1999).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=20480086. PubMed=11027138;
RA Werner R.M., Jiang Y.L., Gordley R.G., Jagadeesh G.J., Ladner J.E.,
RA Xiao G., Tordova M., Gilliland G.L., Stivers J.T.,
RT "Stressing-out DNA? The contribution of serine-phosphodiester
RT interactions in catalysis by uracil DNA glycosylase.";
RL Biochemistry 39:12385-12394(2000).
CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL: J03725; AAA24743.1; -
DR EMBL: D13169; BAA02448.1; -

DR EMBL: D64044; BAA10923.1; -
 DR EMBL: AE000344; AAC75633.1; -
 DR EMBL: D90886; BAA16466.1; -
 DR PIR: A28175; DSECU
 DR PDB: 1EUG; 12-OCT-99.
 DR PDB: 2EUG; 13-OCT-99.
 DR PDB: 3EUG; 13-OCT-99.
 DR PDB: 4EUG; 23-JUL-99.
 DR PDB: 5EUG; 23-JUL-99.
 DR PDB: 1EUI; 22-JUN-99.
 DR PDB: 1UUG; 25-MAR-99.
 DR PDB: 2UUG; 25-MAR-99.
 DR PDB: 1FLZ; 17-JAN-01.
 DR EcoGene: EGI1058; ung.
 DR InterPro: IPR002043; U_DNA_glycosylase.
 DR InterPro: IPR003249; Urac_DNA_glycosyl.
 DR Pfam: PF00315; UNG; 1.
 DR ProDom: PD001589; Urac_DNA_glycosyl; 1.
 DR PROSITE: PS00130; U_DNA_GLYCOSYLASE; 1.
 KW DNA repair; Hydrolase; Glycosidase; 3D-structure; Complete proteome.
 FT INIT_MET 0
 FT ACT_SITE 63 63
 SQ SEQUENCE 228 AA; 25562 MW; E03126BF78F44BB1 CRC64;

Query Match 42.5%; Score 681.5; DB 1; Length 228;
 Best Local Similarity 60.9%; Pred. No. 2.3e-53;
 Matches 131; Conservative 24; Mismatches 59; Indels 1; Gaps 1;

QY 85 TWKRELAAEFKRYFKQLMSFVADER-SRHTVYPPADQVYSTMCIDIDQYKVIILGOD 143
 Db 5 TWHDVLAEEKQDPYFLNLTQTVASERQSGVYIYPPQKDVAFRRFTLGDYKVIILGOD 64
 QY 144 YHGPNOAHGICFSYQKRPVPPPSLVNRYKELCTDIDGFKHFGHDLGMAKOGVLLINAV 203
 Db 65 YHGPNOAHGIAFVSRRPIALPRLNMYKKELENTIPGTRNHGYLSMAKOGVLLINTV 124
 QY 204 LTVRAHQASHKDRGWFETFDVAIKWLSVNRGVEVFLMGSYAHKKGATIDRRKHHVLA 263
 Db 125 LTVRAGQASHASISGFWETFDKYSILINOREGVFLMGSHAKGAIIDKORHHVLA 184
 QY 264 VHPSPLSAHRGFLGCRHFKHFKANGLLKSGTEPIHW 298
 Db 185 PHPSPLSAHRGFGCNHFVLANOMLEORGEPIDW 219

RESULT 4
 UNG_PASMU
 ID UNG_PASMU STANDARD; PRT; 222 AA.
 AC P57807;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG).
 GN UNG OR PM0065
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxId=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PW70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida PM70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE AS
 A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA POLYMERASE OR
 DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AE006042; AAK02149.1; -
 DR InterPro: IPR002043; U_DNA_glycosylase.
 DR InterPro: IPR003249; Urac_DNA_glycosyl.
 DR Pfam: PF00315; UNG; 1.
 DR ProDom: PD001589; Urac_DNA_glycosyl; 1.
 DR PROSITE: PS00130; U_DNA_GLYCOSYLASE; 1.
 KW DNA repair; Hydrolase; Glycosidase; Complete proteome.
 FT ACT_SITE 61 61
 SQ SEQUENCE 222 AA; 24906 MW; 3F9320277FE53367 CRC64;

Query Match 41.5%; Score 665.5; DB 1; Length 222;
 Best Local Similarity 58.1%; Pred. No. 5.9e-52;
 Matches 126; Conservative 29; Mismatches 61; Indels 1; Gaps 1;

QY 84 ETWRELAAEFKRYFKQLMSFVADER-SRHTVYPPADQVYSTMCIDIDQYKVIILGOD 142
 Db 2 KTWKDVIGTEKTPYFKHILIDVQARASGRIVPPQEVESARQLTEFEVAKVIILGOD 61
 QY 143 YHGPNOAHGICFSYQKRPVPPPSLVNRYKELCTDIDGFKHFGHDLGMAKOGVLLINAV 202
 Db 62 YHGPNOAHGIAFVSRRPIALPRLNMYKKELENTIPGTRNHGYLSMAKOGVLLINTV 121
 QY 203 LTVRAHQASHKDRGWFETFDVAIKWLSVNRGVEVFLMGSYAHKKGATIDRRKHHVLA 262
 Db 122 VLTVRAGQASHASISGFWETFDKYSILINOREGVFLMGSHAKGAIIDKORHHVLA 181
 QY 263 AVHPSPLSAHRGFLGCRHFKHFKANGLLKSGTEPINR 299
 Db 182 AHPSPLSAHRGFLGCRHFKHFKANGLLKSGTEPINR 218

RESULT 5
 UNG_HAEIN
 ID UNG_HAEIN STANDARD; PRT; 219 AA.
 AC P43731;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG).
 GN UNG OR HT0018
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxId=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleisemann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Giodok A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Smal K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
 AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
 CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (BY similarity).
 CC -----

```
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC HSPB: U32687; AAC21696.1; -.
CC DR HSSP: P12295; 3EUG.
CC DR TIGR: HT0018; -.
CC DR InterPro: IPR002043; U_DNA_glycosylase.
CC DR InterPro: IPR003249; Urac_DNA_glycosyl.
CC DR Pfam: PF00315; UNG; 1.
CC DR ProDom: PD001589; Urac_DNA_glycosyl; 1.
CC DR ProSITE: PS00130; U_DNA_GLYCOSYLASE; 1.
CC DR DNA_repair; Hydrolase; Glycosidase; Complete proteome.
CC ACT_SITE 61 61 GENERAL BASE (BY SIMILARITY).
CC FT ACT_SITE 219 AA; 24848 MW; C270CC31F7B2E58A CRC64;
CC SQ SEQUENCE

Query Match
Best Local Similarity 39.6%; Score 635.5; DB 1; Length 219;
Matches 124; Conservative 28; Mismatches 64; Indels 1; Gaps 1;

QY 84 EWRRLAAEFKPYFKQLMSFVADER-SRHTVYPPADQVYSTEMCDIODVKVILGQDP 142
DB 2 KMTDVIIGTEKAPRYQHTLQGVHLARASGKTIYRPEQEDFNFAFKYAFEDVKVILGQD 61
QY 143 PYHGPNQAHGCLCSVOKRPPRPSPVLYNYKELCTDIDGFKHFGHGLSGNAKQGVLLNA 202
DB 62 PYHGPNQAHGLASVPEVAIPPSLINTYKELTODISGFQPMSPNGVLYVKAQEGVLLNLT 121
QY 203 VLTVRANQANSHKDRGMEFTDAVIKWLVSNGEVFLMGSAHKKGATIDRRHHVLQ 262
DB 122 VLTVRGMAHSHANLGMERTDKVIAVLNHRKLVFLMGSHAQKGMQIDRTRHLVLT 181
QY 263 AVHPSPLSAHRGFLGCKHFSKANGKLKLSGTEPINMR 299
DB 182 APHPSPLSAHRGFGFCGRHFSKNTSYLSEHGIKRPIDWQ 218

RESULT 6
UNG_DICDI
ID UNG_DICDI STANDARD: PRT; 257 AA.
AC P53766;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG) (Fragment).
GN UGLA.
OS Dicyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
RX NCBI_TaxID=44689;
RN NCBI_TaxID=44689;
RP SEQUENCE FROM N.A.
RA Guyer R.B., Deering R.A.;
RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE.
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
```

```
DR EMBL: U32866; AAA75334.1; -.
DR HSSP: P13051; 1AK2.
DR DicyDb: DD05081; uglA.
DR InterPro: IPR002043; U_DNA_glycosylase.
DR InterPro: IPR003249; Urac_DNA_glycosyl.
DR Pfam: PF00315; UNG; 1.
DR ProDom: PD001589; Urac_DNA_glycosyl; 1.
DR ProSITE: PS00130; U_DNA_GLYCOSYLASE; 1.
DR DNA_repair; Hydrolase; Glycosidase; DNA repair.
CC ACT_SITE 1 1
CC FT ACT_SITE 97 97 POLY-ASN.
CC FT ACT_SITE 257 AA; 29516 MW; 9CFA0898B648759 CRC64;
CC SQ SEQUENCE

Query Match
Best Local Similarity 37.5%; Score 601; DB 1; Length 257;
Matches 117; Conservative 33; Mismatches 63; Indels 4; Gaps 3;

QY 86 WRRLAAEFKPYFKQLMSFVADERS--RHHTVYPPADQVYSTEMCDIODVKVILGQDP 143
DB 39 WKELSGEEFGKAYFKKMITOLNKRYSSEKRPYPPKNEISAFNYAHLLEDVKVILGQDP 98
QY 144 YHGPNQAHGCLCSVOKRPPRPSPVLYNYKELCTDIDGFKH-GRGDLGNAKQGVLLNA 202
DB 99 YHGQGHGLSFSVKKGVSPPSLINTYKELTDIESFKRPLKGLFEPWAKQGVLLNA 158
QY 203 VLTVRANQANSHKDRGMEFTDAVIKWLVSNGEVFLMGSAHKKGATIDRRHHVLQ 262
DB 159 VLTVEETPRSHKDFGNADTDAVLKLSKQDQIVFILMGGFARQKKEKLETKNHLVLK 218
QY 263 AVHPSPLSAHRGFLGCKHFSKANGKLKLSGTEPINMR 299
DB 219 SGHPSPLSI-KHFGCKHFSKNEFLKSGIEIDWK 254

RESULT 7
UNG_PSEDE
ID UNG_PSEDE STANDARD: PRT; 217 AA.
AC P29950;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG) (Fragment).
GN UNG.
OS Pseudomonas denitrificans.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
RX NCBI_TaxID=43306;
RN NCBI_TaxID=43306;
RP SEQUENCE FROM N.A.
RA STRAIN=SC510;
RX MEDLINE=91008976; PubMed=2211521;
RA Crouzet J., Cameron B., Cauchois L., Rigault S., Rouyez M.-C.,
RA Blanche F., Thibaut D., Debussche L.;
RT "Genetic and sequence analysis of an 8.7-kilobase Pseudomonas
RT denitrificans fragment carrying eight genes involved in
RT transformation of precorrin-2 to cobyrinic acid.";
RL J. Bacteriol. 172:5980-5990(1990).
RN [2]
RP IDENTIFICATION, AND CORRECTION OF A PROBABLE FRAMESHIFT.
RA Asplund R.;
RL Unpublished observations (FEB-1993).
CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```



```

CC      AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC      POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE001659; AAD18911.1; -
DR      EMBL; AE002266; AAF38868.1; -
DR      EMBL; AP002547; BAA98981.1; -
DR      HSSP; P12295; 1EUG.
DR      TIGR; CP1099; -
DR      InterPro; IPR002043; U_DNA_glycosylase.
DR      InterPro; IPR003249; Urac_DNA_glycosyl.
DR      Pfam; PF00315; UNG; 1.
DR      ProDom; PD001589; Urac_DNA_glycosyl; 1.
DR      PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
KM      DNA repair; Hydrolyase; Glycosidase; Complete proteome.
FT      ACT_SITE 77
FT      ACT_SITE 77
SQ      SEQUENCE 236 AA; 27006 MW; 1A734AD68B81AA03 CRC64;

```

Query Match 33.7%; Score 539.5; DB 1; Length 236;
 Best Local Similarity 46.2%; Pred. No. 9.9e-41;
 Matches 111; Conservative 34; Mismatches 82; Indels 13; Gaps 4;

```

QY      66 KKAALDKIRAKAPPAGE---TWRELAAEFEKPYFKQLMSFVADERSRHTVYPPADQ 121
DB      2 QNATIDQL-----FVSWQEQPLCWRQQLKEWMSKPYMQQLIFLKEQYKEHTVYPPENC 56
QY      122 VYSTEMCDIQDKVYVILGDDPHGPNQAHGCLCSVOKPYPPPSLVNITKELCTDIDGF 181
DB      57 VFSALRSTPPDQYRVVILGDDPYPGKGQAHGLSFSVPEGORLPPSLINIFRELKTDLGIE 116
QY      182 KHPGHGDLGSGMAKQGVLLNAVLTVRAHQANSHKDRGMETFTDAVIKWLSVNREGVVEL 241
DB      117 NH--KGCLOQSMANGIILLNTVLTVRAGEPFPSHAGKGMELFTDAIVTKLIQERHIIIFVL 174
QY      242 WGSYAHKKGAFI--DRKRHHVLAQVHPSPLSAHKGFLGCKHFSKANGLLKLGTEPIINMR 299
DB      175 WGAARAKKCELLFNRSKHQHAVLSSPHSPPLAHRGFGCSHFSKINLYLNLKLNKPMINWK 234

```

Search completed: August 21, 2002, 10:28:51
 Job time: 233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 21, 2002, 10:24:38 ; Search time 30.24 Seconds

(without alignments)
1721.940 Million cell updates/sec

Title: US-09-758-017a-2

Perfect score: 1603
Sequence: 1 MLFRLGICQRCISSNRVLPG.....SKANGLKLSTEPINRML 301

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SP:Archea:*
- 2: SP:Bacteria:*
- 3: SP:Fungi:*
- 4: SP:Human:*
- 5: SP:Invertebrate:*
- 6: SP:Mammal:*
- 7: SP:Mhc:*
- 8: SP:Organelle:*
- 9: SP:Phage:*
- 10: SP:Plant:*
- 11: SP:Rodent:*
- 12: SP:Virus:*
- 13: SP:Vertebrate:*
- 14: SP:Unclassified:*
- 15: SP:Virus:*
- 16: SP:Bacteriaph:*
- 17: SP:Archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1596	99.6	301	13	Q91983 gadus morhu
2	1421	88.6	301	13	Q91982 gadus morhu
3	1051	65.6	306	11	Q91988 mus musculu
4	1045	65.2	306	11	P97285 mus musculu
5	1018.5	65.3	313	4	Q93028 homo sapien
6	887	55.3	199	11	Q99KR9 mus musculu
7	702.5	43.8	226	16	Q9KPK8 mus musculu
8	648.5	40.5	219	16	Q9JZAI vibrio chol
9	635.5	39.6	247	16	Q9RMH9 neisseria m
10	634.5	39.6	219	16	Q9RHH9 deinococcus
11	632.5	39.5	282	5	Q9JUC4 neisseria m
12	606	37.8	241	16	Q9U221 caenorhabd
13	598.5	37.3	330	10	Q92L05 rhizobium m
14	591	36.9	227	2	Q9LH6 arabidopsis
15	591	36.9	256	16	Q9K320 streptomyce
16	586.5	36.6	231	16	Q9PA28 xyella fas
					Q915H9 pseudomonas

17	569	35.5	221	16	Q9PU40 campylobact
18	559.5	34.9	224	16	Q92C11 listeria in
19	557	34.7	218	16	Q99W30 staphylococ
20	555.5	34.7	298	5	Q9U776 trypanosoma
21	536.5	33.5	221	16	Q92E00 listeria in
22	528.5	33.0	217	16	Q9A072 streptococc
23	474	29.6	314	12	Q93901 equine herp
24	472	29.4	252	12	Q93695 alcelaphine
25	467.5	29.2	256	12	Q993H3 callitrichi
26	466.5	29.1	254	12	Q69273 leporid her
27	463	28.9	253	12	Q99C24 bovine herp
28	458.5	28.6	255	12	Q80892 herpesvirus
29	451	28.1	252	12	Q9YTL9 ateline her
30	446.5	27.9	212	16	Q9PPU2 ureaplasma
31	445.5	27.8	301	12	Q93497 bovine herp
32	445	27.7	273	12	Q91T17 tupaiia herp
33	444.5	27.7	255	12	Q9WR05 macaca mula
34	436	27.2	255	12	P88934 kaposi's sa
35	435.5	27.2	249	12	P88984 murid herpe
36	423	26.4	244	12	Q9YFJ3 human herpe
37	422.5	26.4	300	12	Q9WR44 cercoplithec
38	418.5	26.1	259	12	Q9DM85 rat cytomeg
39	418.5	26.1	333	12	Q9QTE2 gallid herp
40	415	25.9	293	12	Q37930 feline herp
41	409	25.5	255	12	Q69070 human herpe
42	406	25.3	255	12	Q57142 human herpe
43	402.5	25.1	221	16	Q98PV4 mycoplasma
44	389	24.3	305	12	Q9J3N2 human herpe
45	384.5	24.0	313	12	Q9E6R2 turkey herp

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	301 AA.
Q91983	Q91983			
AC	Q91983			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	URACIL-DNA GLYCOSYLASE (EC 3.2.2.3).			
GN	UNG1.			
OS	Gadus morhua (Atlantic cod).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
OC	Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidae; Gadidae;			
OC	Gadus.			
OX	NCBI_TaxID=8049;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER.			
RA	Lanes O., Willassen N.P.:			
RT	"Identification, cloning and expression of nuclear and mitochondrial			
RT	uracil-DNA glycosylase from Atlantic cod(Gadus morhua).			
RT	Characterisation of the cold-active catalytic domain.";			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AJ275971; CAB85707.1; .			
DR	HSSP: P13051; 1AK2			
DR	InterPro: IPR003249; Urac_DNA_glycosyl.			
DR	InterPro: IPR002043; U_DNA_glycosylase.			
DR	Pfam: PF00315; UNG; 1.			
DR	ProDom: PD001589; Urac_DNA_glycosyl; 1.			
DR	PROSITE: PS00130; U_DNA_GLYCOSYLASE; 1.			
KW	Hydrolase; Glycosidase.			
SQ	SEQUENCE 301 AA; 33807 MW; 21866ADNAAD96C5 CRC64;			

Query Match 99.6%; Score 1596; DB 13; Length 301;
Best Local Similarity 99.7%; Pred. No. 1.1e-142;
Matches 300; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Db 1 MLEKLGICQRCISSNRPLPGLLPQLCFSGKMKITPKRKLRSSVNEOKTSSPOLSVLEOLE 60
QY 61 RMAKNKKAALDKIRAKATPPAGFETWRRELAAEFEKPYFKQOLMSFVADERSRHHVYPAD 120
Db 61 RMAKNKKAALDKIRAKATPPAGFETWRRELAAEFEKPYFKQOLMSFVADERSRHHVYPAD 120
QY 121 QVYSTEMCDIDYKVVILGODPYHGNQAHGLCFSVQKRPVPPPSLVNITKELCTDIDG 180
Db 121 QVYSTEMCDIDYKVVILGODPYHGNQAHGLCFSVQKRPVPPPSLVNITKELCTDIDG 180
QY 181 FKPHGHGDLGSMAGQVLLNNAVLTVAHQANSKDRGWEFTDAVTKWLSVNEGVVFL 240
Db 181 FKPHGHGDLGSMAGQVLLNNAVLTVAHQANSKDRGWEFTDAVTKWLSVNEGVVFL 240
QY 241 LMGSYAKKGGATIDRRKHHVLAQVHPSPLSAHRGFLGCKHFSKANGLLKLSGTEPIWMRA 300
Db 241 LMGSYAKKGGATIDRRKHHVLAQVHPSPLSAHRGFLGCKHFSKANGLLKLSGTEPIWMRA 300
QY 301 L 301
Db 301 L 301
```

```
RESULT 2
Q91982 PRELIMINARY; PRT; 301 AA.
ID Q91982
AC Q91982;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE URACIL-DNA GLYCOSYLASE (EC 3.2.2.3).
GN UNG2.
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Lanes O., Willassen N.P.;
RT "Identification, cloning and expression of nuclear and mitochondrial
RT uracil-DNA glycosylase from Atlantic cod(Gadus morhua).";
RT Characterisation of the cold-active catalytic domain.";
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ275972; CAB85708.1; -.
DR HSSP; P13051; 1AKZ.
DR InterPro; IPR003249; Urac_DNA_glycosyl.
DR InterPro; IPR002043; U_DNA_glycosylase.
DR Pfam; PF00315; UNG; 1.
DR ProDom; PD001589; Urac_DNA_glycosyl; 1.
DR ProSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
KM Hydrolase; glycosidase.
SQ SEQUENCE 301 AA; 33853 MW; C5F0BBALDSCA89B CRC64;
```

```
Query Match 88.6%; Score 1421; DB 13; Length 301;
Best Local Similarity 98.9%; Pred. No. 3.8e-126;
Matches 266; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 213 SHKRGWETFDAYIKWLSVNEGVVFLMGSYAKKGGATIDRRKHHVLAQVHPSPLSAH 272
Db 213 SHKRGWETFDAYIKWLSVNEGVVFLMGSYAKKGGATIDRRKHHVLAQVHPSPLSAH 272
QY 273 RGFGLGCKHFSKANGLLKLSGTEPIWMRAL 301
Db 273 RGFGLGCKHFSKANGLLKLSGTEPIWMRAL 301
```

```
RESULT 3
Q9JTW8 PRELIMINARY; PRT; 306 AA.
ID Q9JTW8
AC Q9JTW8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE URACIL-DNA GLYCOSYLASE NUCLEAR ISOFORM.
GN UNG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RX MEDLINE=20330343; Pubmed=10871356;
RA Nilsen H., Steinsbekk K.S., Otterlei M., Slupphaug G., Aas P.A.,
RA Krokan H.E.;
RT "Analysis of uracil-DNA glycosylases from the murine Ung gene reveals
RT differential expression in tissues and in embryonic development and a
RT subcellular sorting pattern that differs from the human homologues.";
RN Nucleic Acids Res. 28:2277-2285(2000).
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF174485; AAF76936.1; -.
DR EMBL; BC011039; AAH11039.1; -.
DR HSSP; P13051; 1AKZ.
DR MGD; MGI:109352; Ung.
DR InterPro; IPR003249; Urac_DNA_glycosyl.
DR InterPro; IPR002043; U_DNA_glycosylase.
DR Pfam; PF00315; UNG; 1.
DR ProDom; PD001589; Urac_DNA_glycosyl; 1.
DR ProSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
SQ SEQUENCE 306 AA; 33926 MW; CE2D5192936CE6EA CRC64;
```

```
Query Match 65.6%; Score 1051; DB 11; Length 306;
Best Local Similarity 72.0%; Pred. No. 3.7e-91;
Matches 188; Conservative 31; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 41 RSSVNEOKTSSPOLSVLEOLERMMAKNKKAALDKIRAKATPPAGFETWRRELAAEFEKPYFK 100
Db 46 KKAAYEDNEOGSPLSAQLVRIQNRKAALLRLAARVVRPGFSGSKQQLCGEGRKRYFV 105
QY 101 QLMSFVADERSRHHVYPADQVYSTEMCDIDYKVVILGODPYHGNQAHGLCFSVQKRP 160
Db 106 KLGMFVAEERNHHKVVYRRPQVFTWTQMDIDYKVVILGODPYHGNQAHGLCFSVQRP 165
QY 161 VPPPSLVNITKELCTDIDGFKPHGHGDLGSMAGQVLLNNAVLTVAHQANSKDRGWE 220
Db 166 VPPPSLENIFKELSTIDGFEVHGHGDLGSMAGQVLLNNAVLTVAHQANSKDRGWE 225
QY 221 TPTDAVTKWLSVNEGVVFLMGSYAKKGGATIDRRKHHVLAQVHPSPLSAHRGFLGCKH 280
Db 226 QFTDAVYVSNQNLNSGLVFLMGSYAOKGSDVDRKHHVLAQVHPSPLSAHRGFLGCRH 285
QY 281 FSKANGLLKLSGTEPIWMRAL 301
Db 286 FSKANGLLKLSGTEPIWMRAL 306
```


RA Strausberg R.: Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC004037; AAH04037.1; -.
 DR HSSP; P13051; 1AKZ.
 DR InterPro; IPR003249; Utrac_DNA_glycosyl.1.
 DR InterPro; IPR002043; U_DNA_glycosylse.
 DR Pfam; PF00315; UMG; 1.
 DR ProDom; PD001589; Utrac_DNA_glycosyl; 1.
 DR PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
 SO SEQUENCE 199 AA; 22526 MW; 26FAD2B1597AC4A CRC64;

Query Match	55.3%;	Score 887;	DB 11;	Length 199;
Best Local Similarity	78.9%;	Pred. No. 6.7e-76;		
Matches 157;	Conservative 20;	Mismatches 22;	Indels 0;	Gaps 0;

[illegible]

RESULT	7			
Q9KPK8		PRELIMINARY;	PRT;	226 AA.
ID	Q9KPK8			
AC	Q9KPK8;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	URACIL-DNA GLYCOSYLASE.			
GN	VC2359.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=666;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EL TOR N16961 / SEROTYPE O1;			
KA	MEDLINE=20406833; PubMed=10953301;			
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,			
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,			
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,			
RA	Ernmoleva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,			
RA	McDonald L., Uitterbeck T., Fleischmann R.D., Niernan W.C., White O.,			
RA	Satzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,			
RA	Fraser C.M.;			
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio			
RT	cholerae.";			
RL	Nature 406:477-483(2000).			
DR	EMBL; AE004306; AAP95502.1; -			
DR	HSSP; P12295; 3E0G.			
DR	TIGR; VC2359; -			
DR	InterPro; IPR003349; UTrac_DNA_glycosyl.			
DR	InterPro; IPR002043; U_DNA_glycosylse.			
DR	Pfam; PF00315; UNG; 1.			
DR	ProDom; PD001589; UTrac_DNA_glycosyl; 1.			
DR	ProSITE; PS00130; U_DNA_glycosylase; 1.			
KW	Complete proteome.			
QO	SEQUENCE 226 AA; 25446 MW; 8C972727FFCD4EA4 CRC64;			

Query Match	43.8%;	Score 702.5;	DB 16;	Length 226;
Best Local Similarity	61.1%;	Pred. No. 2.2e-58;		

	Matches	132:	Conservative	25;	Mismatches	58;	Indels	1;	Gaps	1.
Oy	85	TWRRELAAEFEKPYRKOLMSFVADER-SRHIVYPADQOYSTECMDIODYKVITLGGDP	143							
Dd	6	THMDVIAGNEKOAOAFQOOLTLOEFESORAGKVIYPRPAKDVFNAFRFTERGDKVILLIGDDP	65							
Oy	144	VHGSPNQHLCGSYOKRPVPSPSLYNIKELCTDIDGFKNHGHLSGMAKGCVLLILNAV	203							
Dd	66	YNGPPNQAHGLCSFVLPGVTPPSLVNIKYKLADODIPRGQIPRHGYLSQAOGCVLLNLTV	125							
Oy	204	LTVRAHQANSHDRMEPTTDAVIMWLSVNRGEVFLMLSGSAHNKGCATIDRKRHHVLOA	263							
Dd	126	LVEEGGMASHANHTGWETTTDRVIDALONGHRGLLEFLMGSHAKCKGMIDBQRHHVLYMA	185							
Oy	264	VHPSPLSAHRGFLCGKHFSKANGLIKLGSTEPINMR	299							
Dd	186	PHPSPLSAHRGFLCGKHFSKTNOLLQAOGIAPINMQ	221							
	RESULT	8								
	Q9JZAI	PRELIMINARY;	PRT;	219	AA.					
AC	Q9JZAI;									
DT	01-OCT-2000 (TREMBLrel. 15, Created)									
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)									
DE	01-DEC-2001 (TREMBLrel. 19, last annotation update)									
GN	URACIL-DNA GLYCOSYLASE.									
NC	NMB122.									
OC	Neisseria meningitidis (serogroup B).									
OX	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.									
NCBI_TaxID=491;	[1]									
RN	SEQUENCE FROM N.A.									
RC	STRAIN-MC58 / SEROGROUP B;									
RX	MEDLINE=20175755; Pubmed=10710307;									
RA	Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,									
RA	Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.U.,									
RA	Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,									
RA	Hait D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,									
RA	Cason T., Ciecko A.E., Parksey D.S., Blair E., Clifton E., Clark E.B.,									
RA	Morton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,									
RA	Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,									
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venler J.C.,									
RT	"Complete genome sequence of Neisseria meningitidis serogroup B strain									
RL	MC58."									
SC	Science 287:1809-1815(2000).									
DR	EMBL: AE002470; AAFA1604.1; -									
DR	HSSP: P12395; 3EUG.									
DR	TIGR: NMB1222; --									
DR	InterPro: IPR002106; AA_trna_ligase-II.									
DR	InterPro: IPR003249; Urac_DNA_glycosyl.									
DR	InterPro: IPR002043; U_DNA_glycosylse.									
PFam:	PF00315; UMG: 1.									
Prodrom:	PD001589; Urac_DNA_glycosyl: 1.									
DR	PROSITE: PS00339; AA-trna_ligase-II.2; UNKNOWN_1.									
KM	Complete proteome.									
SQ	SEQUENCE: 219 AA; 2458 MM; 46D6E58FA283B3F2 CRC64;									
	Query Match	40.5%;	Score	648.5;	DB	16;	Length	219;		
	Best Local Similarity	55.8%;	Pred. No.	2.7e-53;						
	Matches 121;	Conservative	35;	Mismatches	60;	Indels	1;	Gaps	1;	
Oy	84	EWWRRELAEPFKPYRKOLMSFVADER-SRHIVYPADQOYSTECMDIODYKVITLGGD	142							
Dd	2	DTMHDLAGEKKOOPFOELILNAVNERLSGGOTIIPPADVFAEFRLTFDYRKAVAILIGD	61							
Oy	143	PYHGNOAQHAGCSFYQKRVPPSPSLYNITYKELCCTIDGFKPHGBHDLSGMAKOGVLLINA	202							
Dd	62	PYHGNGOAHGLAFSVROGIRIPPSLVNIYKLEFDIBEGSFIPAHGCLTAMAEQGVLLNT	121							
Oy	203	VTVAHQANSHKRDGWTFTDAVIMWLSVNREGVFLMLSGSYAHKGCATTIDRRKHHVLO	262							

QY 203 VLTVRHQANSHKDRGNETFTDAVIKMLSVNREGVFFLLMGSAHHKKGATIDRRHHYLQ 262
||||| :||| ||| ||| : : : : : ||| ||| : | : | :

DB 122 VLVYRAGQASHALLGWERFTDVIRQLATHRKHLVFMIMGVYAOQKGLIDSONHLILT 181
QY 263 AVHPSPLSAHGRFLGCKHFKSKANGKLKLGSTEPINMR 299
DB 182 APHPSPLSAVRGFGCGRHFSQANSYLSRHGIDPINMK 218

RESULT 9
Q9RWH9 PRELIMINARY: PRT; 247 AA.
AC 09RWH9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE URACL-DNA N-GLYCOSYLASE.
GN DR0689.
OS Deinococcus radiodurans.
OC Bacteria: Thermus/Deinococcus group: Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Motlat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001926; AAF10269.1; -.
DR HSSP; P13051; 1AK2.
DR TIGR; DR0689; -.
DR InterPro: IPR003249; UraC_DNA_glycosyl.
DR InterPro: IPR002043; UraC_DNA_glycosylase.
DR Pfam; PF00315; UNG; 1.
DR ProDom; PD001589; UraC_DNA_glycosyl; 1.
DR PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
KW Complete proteome.
SQ SEQUENCE 247 AA; 27745 MW; B92EB319AFF4736 CRC64;

Query Match 39.6%; Score 635.5; DB 16; Length 247;
Best Local Similarity 56.1%; Pred. No. 5.4e-52;
Matches 124; Conservative 27; Mismatches 69; Indels 1; Gaps 1;

QY 79 PAQGFGRWRELAEPKPFKOLMSFVADERSHNYVPPADQVYSTECMDIODVYVI 138
DB 20 PAULPEDWQALLPEFSAPFHELTDFLRQERKEYTIYPAPVDVFNALRYTPPLGEVYVLI 79
QY 139 LGODPYHGRQAGLGFCSVQKVPVPPSLVNIYKELCTDIDGFGHFGDLSGMAKGV 198
DB 80 LGODPYHGRQAGLGFCSVQKVPVPPSLVNIYKELCTDIDGFGHFGDLSGMAKGV 139
QY 199 LINAVALTVRAHQANSKRGWETFTDAVIKWLVSNGVYVLLMSGVYAHKKGATIDRRKH 258
DB 140 LINAVALTVRAHQANSKRGWETFTDAVIKWLVSNGVYVLLMSGVYAHKKGATIDRRKH 199
QY 259 HVLQAVHPSPLSAHGRFLGCKHFKSKANGKLKLGSTEPINMR 299
DB 200 VVIESGHPSPLS-EQYFGFTRPSPKTNLEAKGRGPEWQ 239

RESULT 10
Q9JUC4 PRELIMINARY: PRT; 219 AA.
AC 09JUC4;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE PUTATIVE URACL-DNA GLYCOSYLASE (EC 3.2.2.).
GN UNG OR NMA1384.
OS Neisseria meningitidis (serogroup A).
OC Bacteria: Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies K.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
RA Jagers K., Leach S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162755; CAB84630.1; -.
DR HSSP; P12295; 3EUG.
DR InterPro: IPR002106; AA_tRNA_ligase-II.
DR InterPro: IPR003249; UraC_DNA_glycosyl.
DR InterPro: IPR002043; U_DNA_glycosylase.
DR Pfam; PF00315; UNG; 1.
DR ProDom; PD001589; UraC_DNA_glycosyl; 1.
DR PROSITE; PS00339; AA_tRNA_LIGASE_II_2; UNKNOWN_1.
KW Hydrolyase; Glycosidase; Complete proteome.
SQ SEQUENCE 219 AA; 24685 MW; 46C576CC7E2C03F2 CRC64;

Query Match 39.6%; Score 634.5; DB 16; Length 219;
Best Local Similarity 54.8%; Pred. No. 5.7e-52;
Matches 119; Conservative 36; Mismatches 61; Indels 1; Gaps 1;

QY 84 ETWRELAEPKPFKOLMSFVADERSHNYVPPADQVYSTECMDIODVYVILGOD 142
DB 2 DTWMDALGEGKQDPFQELILNAVRQERLSCGIIYPPADVFNARLTAFRVKAIVILGOD 61
QY 143 PYHGPNOAHGLCFQSVQKVPVPPSLVNIYKELCTDIDGFGHFGDLSGMAKGV 202
DB 62 PYHGPNOAHGLCFQSVQKVPVPPSLVNIYKELCTDIDGFGHFGDLSGMAKGV 121
QY 203 VLVYRAGQASHALLGWERFTDVIRQLATHRKHLVFMIMGVYAOQKRLIDSONYVILT 262
DB 122 VLVYRAGQASHALLGWERFTDVIRQLATHRKHLVFMIMGVYAOQKRLIDSONYVILT 181
QY 263 AVHPSPLSAHGRFLGCKHFKSKANGKLKLGSTEPINMR 299
DB 182 APHPSPLSAVRGFGCGRHFSQANSYLSRHGIDPINMK 218

RESULT 11
Q9U221 PRELIMINARY: PRT; 282 AA.
AC 09U221;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Y56A3A.29A. PROTEIN.
GN Y56A3A.29A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for

RT Investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; AL132860; CAB60520.1; -.
 DR HSSP; P13051; IAKZ.
 DR InterPro: IPR003249; UraC_DNA_glycosyl.
 DR InterPro: IPR002043; U_DNA_glycosylse.
 DR Pfam: PF00315; UNG; 1.
 DR ProDom: PD001589; UraC_DNA_glycosyl; 1.
 SQ SEQUENCE 282 AA; 31893 MW; A0814C43077E64BE CRC64;

Query Match 39.5%; Score 632.5; DB 5; Length 282;
 Best Local Similarity 46.8%; Pred. No. 1.2e-51;
 Matches 132; Conservative 46; Mismatches 93; Indels 11; Gaps 5;

QY 30 SKLTKITPKKLRSSNVQKTSPPQLSVQLEERMAKKKALDKTRAKA-TPAGP----- 82
 DB 2 SKTVRIPDMFLKASAKRSASSTENIPKVPAGNENQVKKKLQAPETETILKSL 61
 QY 83 -GETWRRELAEEFEKPYFKOLMSFVADERSR-HTVYPADQVYSTEMCDIODVKVILG 140
 DB 62 TGSWSKLEEEFKKGYISKIEKFLNSEVKKQVPPPPQIETFMLEPFDEISVIIG 121
 QY 141 QDPYHGNQAHGICFSYQKVPVPPPSLVNITYKELCTDIDGFKHPGHDLSGMAKQVLL 200
 DB 122 QDPYHNDQAHGILSFSYQKGVKPPPSLKINITYKELESDEGFKRPDHCNLLGWTROGVML 181
 QY 201 NAVTVRAHQANSHKDGWETFTDAVIKMLSVNRE-GVFFLLGMSYAHKKGATIDRRKHH 259
 DB 182 NAVTVRAHQANSHAKIGWETFTDAVIRITISROSEKPIVFLWGMGFANKEELIDTKHV 241
 QY 260 VLOAVHPSPLSAHSGFLGCKHFSKANGLLKLGSTEPINMAL 301
 DB 242 VIKTAHPSPLSA-RKMWGCKCFKSCNTELENSGRNPIINMADL 282

RESULT 12
 ID 092LU5 PRELIMINARY; PRT; 241 AA.
 AC 092LU5;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DE PROBABLE URACIL-DNA GLYCOSYLASE PROTEIN (EC 3.2.2.2).
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21368234; PubMed=11474104;
 RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F., Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G., Bouty M., Bowser L., Burnmester J., Cadieu E., Capela D., Chain P., Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F., Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gujral M., Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D., Kahn M.L., Kalman S., Keeling D.H., Kiss E., Komp C., Lelaure V., Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B., Remyerger U., Surzycki R., Thebaud P., Vandendol M., Vornheller F.J., Weidner S., Wells S., Wong K., Yeh K.-C., Batut J.;
 RA "The composite genome of the legume symbiont Sinorhizobium meliloti.";
 RL Science 293:668-672(2001).
 DR EMBL; AL591792; CAC47505.1; -.
 KW Hydrolyase; Glycosidase; Complete proteome.
 SQ SEQUENCE 241 AA; 27174 MW; 341ADE4874CE3802 CRC64;

Query Match 37.8%; Score 606; DB 16; Length 241;
 Best Local Similarity 53.5%; Pred. No. 3.2e-49;
 Matches 116; Conservative 32; Mismatches 67; Indels 2; Gaps 2;

QY 84 ETWRRELAEEFEKPYFKOLMSFVADERSR-HTVYPADQVYSTEMCDIODVKVILG 142
 DB 9 ESMKAVLGEERHGYAMELKRLLEEKQGRQJRPGEVEFRALDPLDRVRVILG 68
 QY 143 PYHGNQAHGICFSYQKVPVPPPSLVNITYKELCTDIDGFKHPGHDLSGMAKQVLL 202
 DB 69 PYHNDQAHGICFSYQKGVKPPPSLKINITYKELESDEGFKRPDHCNLLGWTROGV 127
 QY 203 VLVTVRAHQANSHKDGWETFTDAVIKMLSVNREGVFFLLGMSYAHKKGATIDRRKHH 262
 DB 128 VLVTVRAHQANSHKDGWETFTDAVIRAVNDAQVYVMIMGSTYQKRAAIVDRSHLV 187
 QY 263 AVHPSPLSAHSGFLGCKHFSKANGLLKLGSTEPINMR 299
 DB 188 AVHPSPLSAHSGFLGCKHFSKANGLLKLGSTEPIDMR 224

RESULT 13
 ID 09LIH6 PRELIMINARY; PRT; 330 AA.
 AC 09LIH6;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE URACIL-DNA GLYCOSYLASE-LIKE PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the regions of 4,251,695 bp covered by ninety P1, TAC and BAC clones."
 RL DNA Res. 7:217-221(2000).
 DR EMBL; AF001303; BAB02221.1; -.
 DR HSSP; P13051; IAKZ.
 DR InterPro: IPR003249; UraC_DNA_glycosyl.
 DR InterPro: IPR002043; U_DNA_glycosylse.
 DR Pfam: PF00315; UNG; 1.
 DR ProDom: PD001589; UraC_DNA_glycosyl; 1.
 SQ SEQUENCE 330 AA; 36289 MW; 696BC91977309E19 CRC64;

Query Match 37.3%; Score 598.5; DB 10; Length 330;
 Best Local Similarity 45.8%; Pred. No. 2.5e-48;
 Matches 131; Conservative 36; Mismatches 92; Indels 27; Gaps 6;

QY 37 PKLRSSNVQKTSPPQLSVQLEERMAKKKALDKTR-----AKATPAG----- 81
 DB 48 PRVTVTTSVADDSG-LTPEQIARAEEFNFKVAKSRNLAVGSERYTKAKSEGCYPLS 105
 QY 82 ---FGTWRRELAEEFEKPYFKOLMSF-----VADERSRHTVYPADQVYSTEMCDIOD 133
 DB 106 ELVRESWLKALPGEFHKPYAKSLSDLEIREITDSK-PLIIPQHLINALNTTPFDR 164
 QY 134 VKVYVILGDPYHGNQAHGICFSYQKVPVPPPSLVNITYKELCTDIDGFKHPGHDLSGMA 193
 DB 165 VKTVIIGDQYHGNQAHGILSFSYQKGVKPPPSLKINITYKELESDEGFKRPDHCNLLG 223
 QY 194 KQGVLLNAVTVRAHQANSHKDGWETFTDAVIKMLSVNREGVFFLLGMSYAHKKGATID 253
 DB 224 VQGVLLNAVTVRAHQANSHKDGWETFTDAVIQISTQKREGVFLWMGRVQDEKSKLI 283

QY 254 DRKHHVLAQVHPSPSAHSGFLGCKHPSKANGLLKSTGEPINM 299
 Db 284 DATKHILLTAHPSGLSANRCGFPCRHRSKRANQLLEEMGIPIDWQ 329

RESULT 14
 Q9K320 PRELIMINARY; PRT: 227 AA.

AC 09K320
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE URACIL-DNA GLYCOSYLASE (EC 3.2.2.-).
 GN UNG.

OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 ON NCBI_TaxID=1902;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kieser H.M., Denepate D., Eichner A., Cullum J.,
 RA Kinshir H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL359949; CAB95800.1; -

DR HSSP; P13051; 1AKZ.
 DR InterPro: IPR003249; Urag_DNA_glycosyl.
 DR InterPro: IPR002043; U_DNA_glycosylase.
 DR Pfam; PF00315; UNG; 1.
 DR ProDom; PD001589; Urag_DNA_glycosyl; 1.
 DR ProDom; PD00130; U_DNA_GLYCOSYLASE; 1.
 DR PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
 SQ SEQUENCE 227 AA; 25205 MW; 775BFB5A984C2982 CRC64;

Query Match 36.9%; Score 591; DB 2; Length 227;
 Best Local Similarity 53.0%; Pred. No. 7.8e-48;
 Matches 114; Conservative 35; Mismatches 64; Indels 2; Gaps 2;

QY 84 ETWRRELAEEFKPYKOLMSFVADERSRHTVPPADQVYSSTEMCDIDQVYVILGDP 143
 Db 9 ESKREVLGGELDOPYFKELMEFEVEERANGPYPPREVEFAALADANPPFRVKKVLVGDP 68
 QY 144 YHGPNQAHGLCSYOKPVPSPSLVNIYKELCTDIDGFKHGDLSGAKGVLLINAV 203
 Db 69 YHGEGHGLCSYVRGVAVPPSLRIYKEMAEELD-TIPDNGYLPMAEGGVLLINAV 127
 QY 204 LTVRAHQANSHKDRGWEFTTDAVITKLSVNRGVVFLMGSYAHHKGGATIDRRKHHVLA 263
 Db 128 LTVRAHQANSHKDRGWEFTTDAVITKLSVNRGVVFLMGSYAHHKGGATIDRRKHHVLA 263
 QY 264 VHPSPSAHSGFLGCKHPSKANGLLKSTGEPINM 298
 Db 188 AHPSPLSAKK-FFGSRPFTQINEAVAGCGHEPIDW 221

RESULT 15
 Q9PA28 PRELIMINARY; PRT: 256 AA.
 AC 09PA28
 DT 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE URACIL-DNA GLYCOSYLASE.
 GN XF2692.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 ON Xylella.
 ON NCBI_TaxID=2371;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-9A5C;
 RA MEDLINE-2036571; PubMed-10910347;
 RA Simpson A.J.G., Rehnach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
 RA Quaggo R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA da Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,
 RA Valida H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Zelditch J., Zetzel J.C.;

RT "The genome sequence of the plant pathogen xylella fastidiosa";
 RL Nature 406:151-159(2000).
 DR EMBL; AE004075; AAP85489.1; -
 DR HSSP; P12295; 3EUG.
 DR InterPro: IPR003249; Urag_DNA_glycosyl.
 DR InterPro: IPR002043; U_DNA_glycosylase.
 DR Pfam; PF00315; UNG; 1.
 DR ProDom; PD001589; Urag_DNA_glycosyl; 1.
 DR Complete proteome.
 SQ SEQUENCE 256 AA; 28187 MW; 4F21E7001BA1265B CRC64;

Query Match 36.9%; Score 591; DB 16; Length 256;
 Best Local Similarity 53.2%; Pred. No. 9.2e-48;
 Matches 116; Conservative 31; Mismatches 63; Indels 8; Gaps 3;

QY 85 TWRRELAEEFKPYKOLMSFVADERSRHTVPPADQVYSSTEMCDIDQVYVILGDP 140
 Db 24 SMRAHGNWMLRPMEDLSAFL--RAKRVAGSVYVPPGQIPFAAEATPFQVKKAVILG 80
 QY 141 QDPYHGNQAHGLCSYOKPVPSPSLVNIYKELCTDIDGFKHGDLSGAKGVLLINAV 200
 Db 81 QDPYHGNQAHGLCSYVRGVAVPPSLRIYKEMAEELD-TIPDNGYLPMAEGGVLLINAV 127
 QY 201 NAVLTVRAHQANSHKDRGWEFTTDAVITKLSVNRGVVFLMGSYAHHKGGATIDRRKHHV 260
 Db 140 NAVLTVRAHQANSHKDRGWEFTTDAVITKLSVNRGVVFLMGSYAHHKGGATIDRRKHHV 260
 QY 261 LQAVHPSPLSAHSGFLGCKHPSKANGLLKSTGEPINM 298
 Db 200 LKAHPSPLSAHSGFLGCKHPSKANGLLKSTGEPINM 237

Search completed: August 21, 2002, 10:28:33

Thu Aug 22 10:06:21 2002

us-03-758-017a-2.rspt

Job time: 235 sec